DOCUMENT RESUME

ED 127 369 95 TM 005 529

AUTHOR Ekstrom, Ruth B.
TITLE Beginning Teacher Evaluation Study: Phase II, 1973-74, Final Report: Volume IV. Pupil and Teacher Tests.
INSTITUTION California State Commission for Teacher Preparation and Licensing, Sacramento.; Educational Testing Service, Princeton, N.J.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO ETS-PR-76-11
PUB DATE Jul 76
NOTE 95p.; For related documents, see TM 005 524-535

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS Academic Achievement; Achievement Tests; Aptitude Tests; Cognitive Style; Criterion Referenced Tests; Diagnostic Tests; Effective Teaching; Elementary Education; *Elementary School Students; *Elementary School Teachers; Expectation; Films; Grade 2; Grade 5; Interviews; Item Analysis; Knowledge Level; Mathematics; Norm Referenced Tests; Principals; Questionnaires; Reading; Reading Tests; Standardized Tests; Statistical Analysis; Student Characteristics; *Student Testing; Teacher Attitudes; Teacher Behavior; Test Reliability; *Tests; Test Selection

ABSTRACT

The Beginning Teacher Evaluation Study (BTES), Phase II, was a research project on effective teaching behavior--what teachers do that significantly affects what and how pupils learn. The purposes of Phase II were to (1) develop an assessment system for measuring teacher and student behaviors and other factors which could influence each of them and their interrelationships and (2) generate hypotheses about the interrelationships among teacher and pupil behaviors and related factors. Forty-one second grade and 54 fifth grade experienced teachers participated in the study. This volume focuses on the teacher and pupil measures used in the study. Student background measures included demographic data, academic ability, and cognitive style. Student achievement measures included cognitive skills, decoding, reading comprehension, mathematics application, and mathematics attitude. A variety of teacher measures were used--tests of teacher aptitude and cognitive style, teaching and subject matter knowledge, and teacher attitudes toward teaching. Other measures were also included in the study: a diagnostic film test, a teacher questionnaire on the school setting, a teacher background questionnaire, a Likert scale to elicit teacher and principal perceptions of school climate, an instrument to elicit teacher expectations for pupil performance, a pupil survey questionnaire completed by the teacher, a parent questionnaire, a questionnaire for principals on school characteristics, and a structured interview with the principals. (RC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED127369

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
DOROTHY URBAN
ETS
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

TM005 529

PR-76-11

BEGINNING TEACHER EVALUATION STUDY
PHASE II 1973-74

FINAL REPORT

VOLUME IV

PUPIL AND TEACHER TESTS

BY

RUTH B. EKSTROM

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

BEGINNING TEACHER EVALUATION STUDY: PHASE II

FINAL REPORT: VOLUME IV

PUPIL AND TEACHER TESTS

by

Ruth B. Ekstrom

Educational Testing Service
Princeton, New Jersey

A project conducted by Educational Testing Service for the California Commission for Teacher Preparation and Licensing and funded by the National Institute of Education.

Copyright © 1975, 1976 by Educational Testing Service. All rights reserved.

PREFACE

The Beginning Teacher Evaluation Study (BTES) is a long-term project of the California Commission for Teacher Preparation and Licensing. The Commission is responsible for licensing teachers in California and is trying to determine what factors should be considered in this process.

The second phase of the study was conducted by Educational Testing Service for the Commission. Phase II was the hypotheses-generating and instrument-development phase of BTES. ETS had two tasks: (1) to develop an assessment system to measure both teacher and pupil behaviors as well as other factors which might be related to these behaviors; and (2) to generate hypotheses about the interrelationships between teacher and pupil behaviors and related factors.

The study was conducted in 43 schools in eight districts throughout the state of California. A total of 41 second grade teachers and 54 fifth grade teachers participated in the project during Phase II.

The final report for Phase II consists of several volumes. Volume I describes the design and rationale for the experimental design and data analysis procedures and includes the major findings of Phase II. Volume II describes the conduct of the field study and the sample of participants.

Because of the complex nature of Phase II, a variety of techniques was used to measure teacher and pupil behaviors. They are described in Volumes III, IV, and V. Results are also included in these volumes.

Volume III describes the observation systems in detail and is available in three separately bound sections. The first section, Volume III.1., describes the behavior recording observation system used in the project--APPLE (Anectdotal Process for Promoting the Learning Experience). Volume III.2. describes the category system used to observe classroom activities--

RAMOS (Reading and Mathematics Observation System). The third section of this volume, III.3., covers the videotaping of instructional activities during reading and mathematics.

Volume IV concerns other aspects of the measurement system and covers both the pupil and teacher test batteries.

The fifth volume covers a series of small studies done as part of Phase II. Volume V.1. looks at teacher aptitudes as related to teacher behaviors. Volume V.2. is concerned with the relationship between teacher expectations and pupil performance. Volume V.3. reviews performance of pupils in the BTES teachers' classrooms for two years prior to Phase II, the historical test data. Volume V.4. discusses the Diagnostic Film Test, a device designed to assess teachers' skills in diagnosing reading problems and prescribing corrective action. Volume V.5. summarizes the results of work diaries completed by the teachers on their reading and mathematics instructional program.

Information on the availability of these volumes can be obtained from:

Dr. Frederick J. McDonald
Educational Studies
Educational Testing Service
Princeton, NJ 08540

Information on other phases of BTES can be obtained from:

California Commission for Teacher
Preparation and Licensing
1020 O Street
Sacramento, CA 95814

ACKNOWLEDGEMENTS

The author wishes to acknowledge the contributions of many to this study: Meredith Stone, Leta Davis, Paula Held, and Evelyn Manley for the preparation of the data for analysis; Barbara Foltin and William Allen for the analyses of the data; Gina Wilson, Virginia Vandagriff, and J. Richard Harsh for coordinating the activities in Southern California; Mary Corder, Helga Grabske, Pamela Roelfs, Alan Gold, Linda Cole, and Shirley Loret for their efforts in gathering the teacher, principal, and pupil data; Florence Noe and Freda Davis for collating and packaging the tests and reports; Sharon Tucker, Janice Jackson, and Joanne Farr for manuscript and test preparation; John Stevenson and Elvin Ball for printing; and, Jean Gutterm[illegible] for her unfailing support and cooperation as the coordinator of all t[illegible] support services in the Berkeley Office.

The author also wishes to acknowledge the help of the personnel in each participating district for identifying and recruiting the substitute teachers who worked with us to administer the pupil tests. A special acknowledgement goes to the Research Advisory Board and the Evaluation of Teaching Committee of the California Commission for Teacher Preparation and Licensing for their suggestions.

The most important acknowledgement of all is for the efforts of the pupils, teachers, and principals who made this study possible.

Ruth B. Ekstrom

TABLE OF CONTENTS

	Page
Preface	1
Acknowledgements	iii
Pupil and Teacher Tests	1
Pupil Measures	4
Test Type	4
Test Selection	6
Test Development	13
Test Review, Try-Out and Revision	25
Testing Procedures	26
Additional Test Modification	27
Scoring	29
Item Analysis	30
Descriptive Statistics	31
Relationship Between the Tests and Instruction	40
Teachers Measures	43
Relevance of Teacher Aptitude	43
Selecting Teacher Aptitude Dimensions	45
Tests Selected for Aptitude Measurement	47
Teacher Knowledge	48
Implications	52
Cognitive Style	53
Pilot Testing and Revisions	54
Testing Procedures	54

TABLE OF CONTENTS (continued)

Page

Scoring and Data Reduction 56

Descriptive Statistics 57

Other Measures 62

Diagnostic Film Test 62

Teacher Questionnaires 63

Likert Questionnaire 67

Teacher's Expectations of Pupils 70

Pupil Information from the Teacher 71

Parent Questionnaire 72

Principal Questionnaire and Interview 73

References 79

Appendix 81

LIST OF TABLES

Table		Page
1	Pupil Test Battery - Grade 2	23
2	Pupil Test Battery - Grade 5	24
3	Schedule for Test Administration	28
4	Summary of Item Analysis Data	32
5	Means ($\bar{X}$) and Standard Deviations (SD) for Grade 2 for Pupils with both Fall and Spring Scores	33
6	Reliability of Grade 2 Pupil Tests	34
7	Intercorrelations for Grade 2 Test Battery	35
8	Means ($\bar{X}$) and Standard Deviations (SD) for Grade 5 for Pupils with both Fall and Spring Scores	36
9	Reliability of Grade 5 Pupil Tests	37
10	Intercorrelations for Grade 5 Test Battery	38
11	Correlations Between Amount of Time and Frequency of Teaching Certain Reading Skills and Pupil Test Mean Gains	41
12	Correlations Between Amount of Time and Frequency of Teaching Certain Mathematics Skills and Pupil Test Gains	42
13	Teacher Test Scores	50
14	Test Statistics for Fall Teacher Measures	58
15	Test Statistics for Spring Teacher Measures	59
16	Teacher Test Intercorrelations	61
17	Selected Fall Teacher Questionnaire Responses	64
18	Selected Teacher School Climate Questionnaire Responses	68
19	Pupil Survey Instrument	71
20	Parent Questionnaire	74

LIST OF FIGURES

Figure		Page
1	A structural model of the domain of variables influencing teaching performances and children's learning	3
2	Dimensions of reading comprehension	15

Pupil and Teacher Tests

The major emphases of Phase II of the Beginning Teacher Evaluation Study were to generate hypotheses about the relationship between teaching behavior and pupil learning and to develop instruments which would aid in the study of this relationship. These two areas are, of course, closely related. Without some initial hypotheses, it is impossible to develop or select appropriate test instruments. The information obtained from carefully designed measurement instruments helps, in turn, to modify the initial hypotheses and to facilitate understanding of the processes being investigated.

This volume will focus on the instrument development aspect of the Beginning Teacher Evaluation Study. The first section will deal with the pupil tests; the second section with the teacher tests. Observation instruments are covered in Volumes III.1. and III.2.

In each section, the initial suggestions for instrumentation, as described in such sources as the Design Report, the Planning Conference Report, and the Request for Proposal, will be reviewed. The subsequent decision-making process in considering these recommendations, both in terms of the overall design of the study and also in terms of existing test reviews, will be described. The process of developing the measures will be outlined. Finally, the characteristics of each measure used in the study will be described.

This volume is intended as an elaboration on the material presented in Chapter 5 of Volume I, The Measurement System. It provides more detail on the conceptual background, the scoring procedures, and the data outlined on each measure.

As has been explained in Volume I, the design of this study was based on a structural model (shown in Figure 1) of the variables influencing teaching behavior and pupil learning.

Pupil learning was measured by some of the pupil tests discussed in this volume. Except for student behavior, which was measured with an observational instrument, all of the other pupil variables which were hypothesized to have direct impact on student learning are also measured by other pupil tests. These variables include pupil aptitude, cognitive style, attitudes, and expectations.

Teaching behavior was measured primarily with observational techniques. However, many of the variables hypothesized to affect this performance directly, including teacher aptitude, attitudes, knowledge of the subject, and knowledge of teaching, were measured by teacher tests.

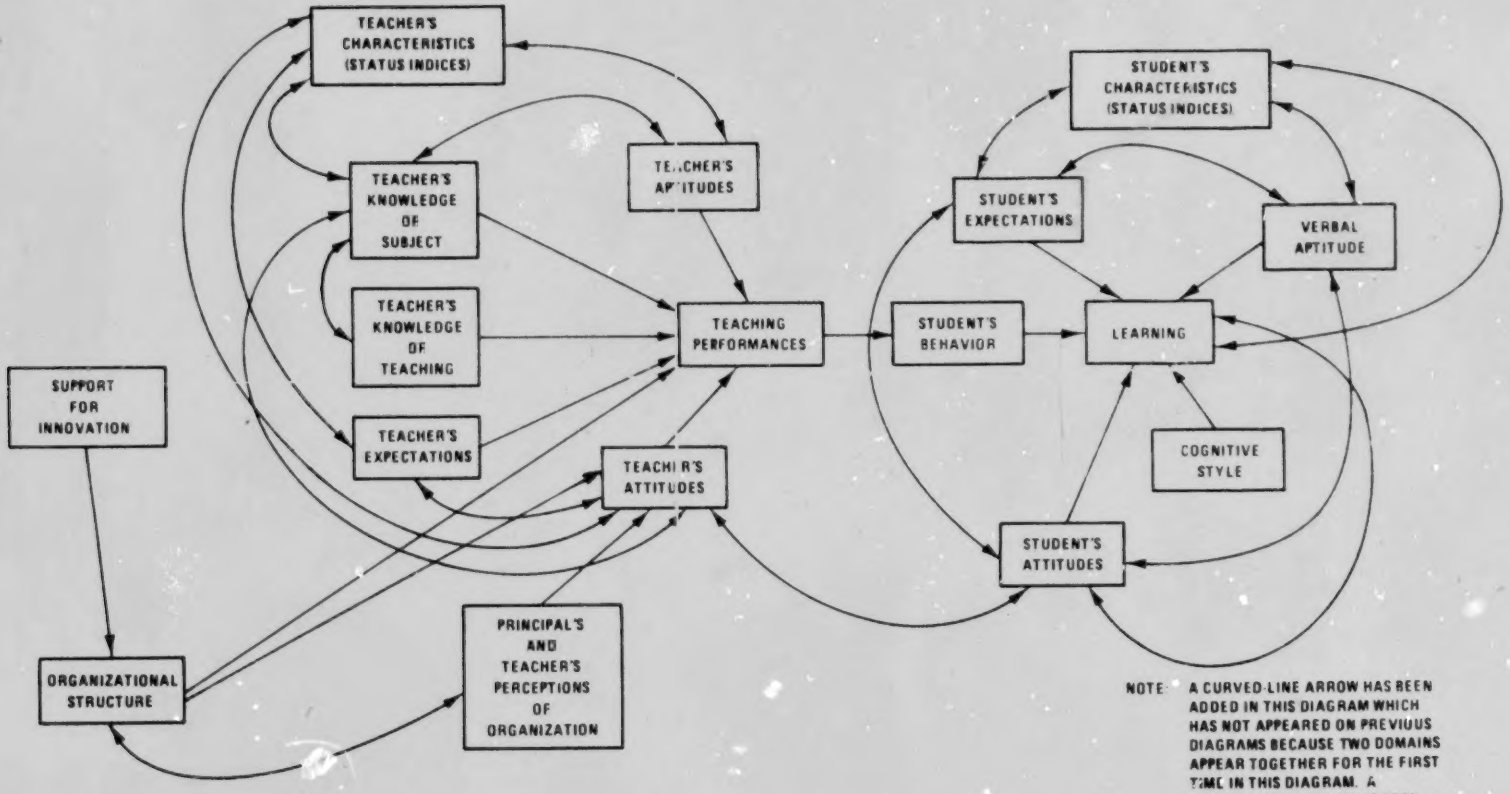


Figure 1. A Structural Model of the Domain of Variables Influencing Teaching Performances and Children's Learning

NOTE: A CURVED-LINE ARROW HAS BEEN ADDED IN THIS DIAGRAM WHICH HAS NOT APPEARED ON PREVIUUS DIAGRAMS BECAUSE TWO DOMAINS APPEAR TOGETHER FOR THE FIRST TIME IN THIS DIAGRAM. A CURVED-LINE ARROW HAS BEEN DRAWN BETWEEN TEACHER'S ATTITUDES AND STUDENT ATTITUDES. STUDENT BEHAVIOR IS REPRESENTED FOR THE FIRST TIME IN THIS DIAGRAM.

Pupil Measures

The research design report suggested two main classes of pupil measures: (1) student background and (2) student achievement. Each of these classes was further subdivided. For example, student background measures included demographic data, academic ability, and student learning style. The student achievement measures included cognitive skills, decoding, reading comprehension, reading application, reading attitude, mathematics concepts and computation, mathematics application, and mathematics attitude.

Test Type

The question of whether to use norm-based or criterion-based measures was first raised in the Planning Conference Report (Commission for Teacher Preparation and Licensing, 1973). The Pupil Progress Group suggested in that report:

> "that the instruments employed to assess pupil performance be criterion-based, not norm-based. That is to say, these measures should be inherently creditable as direct indicators of what pupils are capable of doing as a result of instruction. For purposes of this study, there is no reason that scores on performance measures be related to norms, whether national norms or local norms. In fact, there are probably good reasons not to present the data in terms of norms in order to avoid entirely the possibility of school or district comparison."

The Research Design Report (National Education Program Associates, 1973), however, suggested that both standardized norm-referenced and criterion-referenced tests be used to measure student achievement.

In the literature review prepared by the National Education Program Associates (1973), it was pointed out that standardized achievement tests have been widely used as the criterion for pupil

academic growth. According to this review there have been no studies, to date, which have used criterion-referenced tests when studying the relationship between teacher behavior and student achievement.

As was pointed out in the ETS Literature Review (Ekstrom et al, 1973), "The use of standardized or locally developed tests to measure gains in student achievement is a hazardous procedure." Four major problems related to the selection or development of pupil achievement tests were discussed: (1) difficulty in matching the test to the ability level of the pupils, (2) difficulty in matching the test to the school curricula, (3) difficulty in matching the test to the educational objectives of the teachers, even when curriculum is held constant, and, (4) content validity of the test.

Since, however, there was no basis in data to substantiate the theoretical arguments favoring criterion-referenced tests, it was decided to include both norm-referenced and criterion-referenced pupil achievement tests in this study. Because this was an instrument development and hypothesis generating year for BTES, it seemed to be a most appropriate opportunity to compare the strengths and weaknesses of these two approaches. Additionally, because other educational research work has been criticized for using either one or the other of these approaches exclusively, it seemed wise to use both in order to avoid similar attacks on the design of this study.

Test Selection

The next step in the measure selection and development process was to review the tests which had been suggested in the RFP, to compare these with other similar tests, and to consider the available alternatives in constructing the test battery.

The following measures were recommended in the RFP:

I. Student Background

A. Academic Ability
Short Form Test of Academic Attitudes [sic], Level 1 and 3

B. Student Learning Style
Learning Style Indicator (student form)

II. Student Achievement

A. Reading-cognitive skills, decoding and comprehension

1. Norm-referenced
California Achievement Test, Levels 1 and 3

2. Criterion-referenced
Prescriptive Reading Inventory or Criterion Reading

B. Reading application
National Assessment Project and/or measures to be developed in conjunction with mathematics application

C. Reading attitude
Instrument developed by George Huff

D. Mathematics concepts and computation

1. Norm-referenced
California Achievement Test, Levels 1 and 3

2. Criterion-referenced
Prescriptive Mathematics Inventory

E. Mathematics application
National Assessment Project and/or measures to be developed in conjunction with reading application

F. Mathematics attitude
Hoyt Scale for Measuring Arithmetic Attitude or instrument to be developed based on the Q-sort model developed by George Huff

A major problem in reaching the final decision on a measure of academic ability was whether to treat this as a series of skills (e.g., vocabulary, memory, reasoning, etc.) or as a unitary trait (such as Spearman's "g"). In particular, the question arose as to whether vocabulary should be considered as a measure of reading achievement, as a measure of aptitude, or both.

It was decided to treat aptitude as a multi-factor concept. Such a treatment would permit the analysis of the relationships between specific aptitudes and achievement and to determine if certain aptitude patterns were related to greater achievement in certain subjects. Additionally, if both teacher and pupil aptitude were treated as separate multi-factor entities, it would be possible to look at the interaction between these aptitude patterns and to ascertain which, if any, enhance pupil learning. Vocabulary was treated as an aptitude component.

The two strongest candidates for selection as the student aptitude measure in this study, based on evidence from test reviews such as Buros' Mental Measurement Yearbooks, CSE Elementary School Test Evaluations, and the like, were the Short Form Test of Academic Aptitudes (CTB/McGraw-Hill) and the Otis-Lennon Mental Ability Test (Harcourt, Brace, Jovanovich).

The Otis-Lennon Test requires no reading at the second grade level. Items are grouped into three sections according to type:

inductive reasoning, analogies, and a combination of following directions and picture vocabulary. The test yields part scores and a total raw score, from which IQ, percentile rank, stanine, and mental age equivalents can be computed. At the fifth grade level many of the items require reading. The content includes analogies, vocabulary, computation, etc., with the items spiraled by type through the test. A single raw score for the total test is computed. The differences in measuring aptitude at the two grade levels to be used in this project were considered to be a major problem and eliminated this test from further consideration.

The Short Form Test of Academic Aptitude has four separate subtests: verbal ability, analogytype reasoning, classification-type reasoning, and memory. Scores include both language and nonlanguage aptitude. In reviewing this test with one of the project consultants it was noted that some items in the vocabulary section might present a problem to children using a black dialect and that several of the analogies had a rural bias. Despite these minor difficulties, this test has received generally good reviews and it did not seem possible to find an aptitude measure which could be considered to be significantly better.

Brief consideration was also given to using the tests from the Lesser Study (Lesser et al, 1964) or other tests similar to them as aptitude measures. These tests, which have been used to assess the effects of social class and ethnicity on intellectual performance, include measures of verbal, reasoning, number, and spatial abilities.

However, since the tests were designed for use with six- and seven-year old children, considerable developmental work would have been necessary to produce a comparable test battery for the fifth graders. Such an effort was beyond the scope of this project, given the time and budget constraints.

The learning style measure recommended in the RFP is an experimental measure developed by a group working in the Attleboro, Massachusetts public schools. It attempts to ascertain the pupil's preferred mode of learning in three areas: (1) visual vs. auditory, (2) spatial vs. temporal, and (3) impulsive vs. reflective. However, there was no supporting research evidence for this instrument.

Another closely related area, cognitive style, does have a considerable body of research relating it to classroom learning and to school success. Witkin (1973) has shown that the congitive style known as field dependence-independence, affects pupils' academic development at every educational level and is important in understanding how pupils learn, how teachers teach, and the classroom interaction of teachers and pupils. The research in this area suggests that a match between pupil and teacher congitive style may facilitate pupil learning. There is also evidence that field-independent pupils are more likely to do well in mathematics than field-dependent pupils. It was decided that the inclusion of a congitive style measure, rather than a learning style measure, in this reasearch design would allow the work on the BTES project to be related to a large body of existing research. In addition, the Group Embedded Figures Test, a widely used measure of congitive style, has been used extensively so that there is substantial data

about the psychometric properties of the measure for older children and adults. The existing test for the second grade level, the Children's Embedded Figures Test, is individually administered and had to be revised into a group version for this project. The project staff was fortunate in receiving the cooperation of both Dr. Witkin, the author, and the publisher, Consulting Psychologists Press, in making this adaptation.

The final test selections for measuring pupil aptitude and cognitive style were the Short Form Test of Academic Aptitudes, Levels 1 and 3, and the Group Embedded Figures Test. As noted in Volume I, both of these are treated as trait variables.

There were three possible strategies in selecting or developing the pupil achievement tests: (1) to use a single achievement test for assessing both reading and mathematics achievement, (2) to assemble tests or subtests from various achievement tests in order to obtain the best possible measure in each area, or (3) to develop new tests from either new material or existing item pools so that every item of the tests is maximally appropriate. Option #1 was considered the easiest but least satisfactory; it would have necessitated choosing one standardized norm-referenced test as the sole criteria of pupil growth in this study. As mentioned earlier, it seemed wiser to use both criterion-referenced and norm-referenced measures. Option #2 had the drawback of the cost and difficulty in obtaining permission from other test publishers to use their material and Option #3 entailed the most work. It was decided, after consultation with the Commission, that a combination of #2 and #3 was the optimum strategy.

The following test choices were then made: To measure decoding, (1) the Stanford Achievement Test, Word Study Skills Subtest (Harcourt, Brace, Jovanovich), as a norm-referenced measure and, (2) tests of letter-sound correspondence and root words to be developed by a project consultant, Dr. Robert Calfee of Stanford University. The Stanford Achievement Test was the measure of decoding skills receiving the best reviews for use at the two grade levels being considered. A minor problem existed in using this test since there is a form change in the middle of each grade being tested. It was recommended that the same form be used for pre- and post-testing and that the Commission review the test data before beginning Phase III to ascertain the degree to which pupils "top-out" on these forms and to determine the appropriateness of the difficulty level. Although the California Achievement Test had been recommended in the RFP as a norm-referenced measure of both decoding and comprehension, it does not include any subtest to measure decoding ability. The Prescriptive Reading Inventory has a relatively high proportion of items which could be considered to be ethnically biased because more than one option might be considered to be correct by children using non-standard English. For example, out of the first 83 items on the Red Book (Grade 2) version of this test, at least 16 would present problems for children using a black dialect.

To measure comprehension, (1) the California Achievement Test, Reading Comprehension Subtest omitting the reference skills section, as the norm-referenced measure and, (2) a criterion-referenced test of reading comprehension developed from materials in the ETS test item files. Although the California Achievement Test was recommended

in the RFP as a norm-referenced test of reading comprehension, there is a problem of confounding when using these norms since they include both reference skills and comprehension in a single score. Since it would not be necessary in the analysis to relate pupil raw scores to the norms, it was decided to use this test. The Prescriptive Reading Inventory recommended in the RFP was excluded from further consideration when it was discovered that it would require about three hours of testing time to administer this one instrument. At the fifth grade level, the Comprehensive Tests of Basic Skills, Level II, Reading Comprehension Subtest was a strong candidate competing with the California Achievement Test as a norm-referenced measure of reading comprehension but it had no major advantage over the CAT.

To measure the application of reading skills, the RFP suggested using items from the National Assessment Project. This proved to be an excellent idea. However, there was a minor problem encountered in using the National Assessment Materials since there were only a limited number of reading items suitable for measuring application of reading skills. Therefore, additional items were developed for the reading application tests.

To measure attitude toward reading, items from the attitude survey used in the USOE Study of Compensatory Reading were selected. The Huff Q-sort materials recommended in the RFP would have required individual administration. In addition, the technique is probably too complicated to be understood by many elementary school age children and requires too much reading to be used with second grade pupils. The strongest candidate for a norm-referenced measure of attitude toward reading

was the San Diego County Reading Attitude Inventory. This was the only reading attitude measure found which had both normative and validity information. However, it was decided that there was no particular need to add a norm-referenced attitude scale to the Compensatory Reading Scale.

To measure mathematics concepts and computation, the RFP again suggested the California Achievement Tests. Although there were several other possible tests that were strong alternatives, including the Arithmetic Concepts Subtest of the SRA Achievement Series (Science Research Associates) and the Arithmetic Operations portion of the Comprehensive Test of Basic Skills, (CTB/McGraw-Hill), the California Achievement Test appeared to be the best test combining both of these areas.

To measure mathematics application, the RFP had suggested using National Assessment Mathematics items. Unfortunately, these items have not yet been released to the public so it was necessary to develop a mathematics application test from material in the ETS files.

The Huff Q-sort technique, suggested in the RFP as means of assessing attitude toward mathematics as well as reading, was rejected for the reasons outlined earlier. It was decided to develop a mathematics attitude scale, paralleling the reading attitude scale.

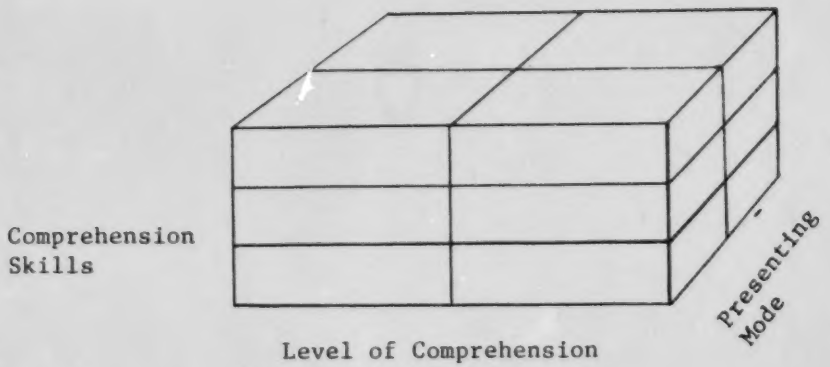
Test Development

As was described above, test development work was needed to construct all or part of measures for decoding, reading achievement (comprehension), reading application, mathematics application, and

attitude toward mathematics. Additionally, test development work was needed to make a group administrable version of the Children's Embedded Figures Test. A Pupil Survey was developed to assess pupils' expectancies for their own success and achievement in reading and mathematics and also their expectancies for their classmates.

As a first step in developing the reading tests, the California Basic Reading Objectives were reviewed. On the basis of this review, it was decided to include the following subtests in the criterion-referenced portion of the second grade decoding test: (1) picture-initial sound correspondence, (2) sight words, (3) root words, and (4) word sounds. The criterion-referenced subtests at the fifth grade level were: (1) syllabification, (2) root words, and (3) phonetic analysis.

Before developing the reading comprehension and reading application tests, a three-dimensional model of the relevant sub-skills was developed (Figure 2). The dimensions are: (1) level of comprehension - literal versus inferential, (2) comprehension skill - main idea, relations, details, and (3) presenting mode - literary versus applied.



Level of Comprehension: Literal
Inferential

Comprehension Skills: Main Idea
Relations
Details

Presenting Mode: Literary
Applied

Figure 2. Dimensions of reading comprehension.

Definitions were developed for each category. Literal questions are those which can be answered solely on the basis of the material contained in the stimulus or reading selection. Inferential questions require either information beyond that presented to the subject or reasoning from the presented material. Detail questions are those which can be answered by referring to one sentence or one pictorial stimulus. Relations questions are those which can be answered by combining the information in two or three sentences or pictorial stimuli. Main ideas are those which involve the synthesis of an entire paragraph, selection, or set of stimuli. Literary questions involve prose selections from both fiction and non-fiction. Application questions are based on pictorial stimuli or written materials drawn from real-life situations.

A faceted design was utilized for both the Reading Achievement (Comprehension) and Reading Application tests. Each test contains six main facets derived from combining two levels of comprehension with the three comprehension skills. Within each of these six facets there are three questions, one which is relatively easy, one of medium difficulty, and one which is of more than average difficulty.

As was mentioned earlier, the Reading Achievement Test items were selected primarily from the ETS item pool. The majority of second grade items was chosen form the Cooperative Primary Tests and the fifth grade items came chiefly from the Sequential Tests of Educational Progress. The test items were classified by three researchers, all of whom are experienced in elementary school reading. Rater agreement was high for the three comprehension

skills (main idea, relations, and details) but somewhat less consistent for level of comprehension (literal vs. inferential). The levels of comprehension appear to be perceived by the reviewers more nearly as a continuous scale than as a dichotomy. An attempt to use three levels of comprehension--literal, paraphrased, and inferential--failed to increase rater agreement significantly.

The Reading Application Test items were drawn primarily from the National Assessment Reading items, but items from other sources as well as new items developed for this project were also included. It had been hoped that it would be possible to obtain permission to use the items developed by ETS for the Adult Functional Reading Study, but these items have not yet been released by the government. However, they proved to be good prototypes for developing measures of reading application. The application items were classified into the same six facets as were the comprehension items thus allowing for analysis of pupil growth across presentation mode by level of comprehension and/or by comprehension skills.

Using test items from existing item pools presents several difficulties. First, it is extremely unwise to rely on any pre-existing item classifications. The categories used and the entire conceptual framework within which the subject matter area is viewed may differ greatly from one set of items to another and have little relationship to the categories appropriate for the present research. Secondly, most item pools do not sample uniformly across all categories of items so that, for a typical set of reading items, the supply of relationship-type items, such as sequence or cause-and-effect, are quickly exhausted while there are still many items

based on details. A related problem is caused when "main idea" is included as a reading item category. This tends to force the test design toward a new reading selection for almost every main idea question asked and, in turn, greatly increases the number of selections which must be included and the amount of time allowed for their reading. Thirdly, most item pools classify items only on a single content dimension but reading content areas are usually better conceptualized as multi-dimensional, as are, for example, the California Basic Reading Objectives. This over-simplification of item content in existing classifications can be confusing or even misleading; at best, it leaves the item reviewer with a substantial amount of classification still to be done.

The Reading Attitude Test items were selected from a longer scale developed for the Compensatory Reading project. Items were chosen on the basis of their psychometric properties in the earlier study. The Planning Conference Report had suggested that unobtrusive measures of attitude toward reading might be obtained from observational records by looking at such indicators as pupils selecting reading when given the opportunity to choose an activity, bringing reading material to class, etc. It would be interesting to ascertain if the classroom observation techniques have yielded enough such behaviors to allow the development of a scale of indicators of positive attitude toward reading. If such a scale could be developed, comparison of scores on it and the paper-and-pencil attitude measure would help to provide validity information for the later instrument.

Because it seemed desirable to ascertain the degree to which the particular test form might account for some of the results,

alternate forms were designed for most of the reading tests. There were five forms of the criterion-referenced word sounds of phonetic analysis portion of the decoding tests in the fall test administration. This was reduced to four forms for the spring administration, primarily by eliminating the poorest items. Two forms of the Reading Achievement Tests and of the Reading Application Tests were administered in the spring. Each of these alternative forms included approximately half of the items from the single fall form, with the remainder being parallel new items.

In developing parallel items for alternative forms of the Reading Application Tests, efforts were made to have questions as similar as possible. For example, an item requiring the pupil to read the wrappers on several types of bubble gum to select the type of gum which would "stay sweet a long time" was paralleled by writing an item requiring the wrappers on several types of candy bars to select "the candy bar with nuts in it". The stimuli for parallel items were similar in both format (written vs. pictorial) and content topics. The type of response required in parallel items (mark the picture, mark the sign, etc.) was also the same.

In selecting parallel reading material and items for the Reading Achievement Tests, the same sources (the ETS item file and National Assessment materials) were used for all items. While it was impossible to have selections dealing with comparable topics, both forms contain non-fiction and fiction materials. As will be seen later in Tables 5 and 8, Form 1 was slightly more difficult than Form 2. This may be related to the higher proportion of non-fiction reading material in Form 2.

The previously used items and materials were assigned to forms to balance difficulty as much as possible based on item data from previous administrations at the same grade levels. The new reading materials and items were then chosen for similar levels of difficulty and to complete the faceted content design. The application items were written to match the needed design facets.

The development of the Mathematics Application Test presented a more complex problem than did the reading tests. The National Assessment Mathematics items had not yet been released to the public so this source, which was recommended in the RFP, was not available. The Planning Conference Report had suggested that a task such as requiring the child to read directions to assemble a toy might be used for this purpose as well as for testing application of reading skills. The major conceptual difficulty with this suggestion is that it becomes nearly impossible to determine whether the student who fails the task does so because of inability to apply reading skills, difficulty with non-mathematical reasoning skills, or inability to apply mathematics skills. In addition, such a task would have required individual testing of all pupils, an impossible constraint given the time and budget limitations of this project. In reviewing what other researchers had used as measures of the application of mathematics skills, it became apparent that mathematics application is often equated with solving verbally stated problems. Consultation with the Commission indicated that such a conceptualization of the application of mathematics skills was considered too restrictive for this study and again introduced the confounding of reading and mathematic skills. It was finally decided to develop a mathematics

application test with three main facets: (1) the application of mathematical concepts, (2) the application of computational skills, and (3) using mathematics in real-life situations. The application of concepts included understanding of mathematical terminology and visual representation of mathematical processes. The application of computational skills included selecting the appropriate numerical representation for an orally presented problem and choosing the appropriate arithmetic process to solve a problem. The real-life situations included costs of items on menus or in advertisements, the construction of simple objects, and simple counting games. Each level of this test had 18 items, six for each of the three facets. Some of the items were selected from the ETS item file, mostly from the Cooperative Primary Tests or the Sequential Tests of Educational Progress, while others were written specifically for this project.

The Mathematics Attitude Test was developed especially for this project. The items were written to parallel those selected for the Reading Attitude scales.

Development of a group administrable form of the Children's Embedded Figures Test was primarily a task of modifying the instructions for group usage and modifying the item format to reduce pupil confusion about the two different simple figures. The modification also involved printing the test in blue ink, rather than black ink, and using various types of "zip-a-tone" shading rather than colors. The use of blue ink rather than black makes it simpler to locate pupil pencil marks tracing the embedded simple figure. Substitution of various shadings for color printing lowered production costs significantly without eliminating the additional embeddedness the color provided.

Validation of this adapted instrument as a measure of cognitive style was not possible within the limitations of this study. However, there is no reason to believe that the changes made from an individual to a group measure should seriously affect the validity of this test.

Pupils' expectations for their own success in reading and mathematics were expected to be closely related to their attitudes toward these subjects. A measure to assess such expectations, the Pupil Survey, was designed by Dr. Marlaine Lockheed of ETS. This measure included questions such as asking the students to indicate which in a series of increasing difficulty of reading and mathematics questions they felt they could do correctly.

The test batteries are described in Tables 1 and 2. The code numbers are used in tables elsewhere in this report. In addition, sample test items are given in the Appendix.

The pupil tests probably measure both direct and incidental learning. It was hoped that the criterion-referenced measures, which were designed to be closely related to State instructional objectives, would be more sensitive to direct learning than the norm-referenced measures.

Table 1

Pupil Test Battery - Grade 2

Test	Title	Form	Number of items	Timing
201	CAT Reading Comprehension*			
	Fall - Level 1, Form A		24	16 minutes
	Spring - Level 2, Form A		35	24 minutes
202	CAT Mathematics Concepts*			
	Fall - Level 1, Form A		32	7 minutes
	Spring - Level 2, Form A		30	8 minutes
203	CAT Mathematics Computation*			
	Fall - Level 1, Form A		32	14 minutes
	Spring - Level 2, Form A	(addition and subtraction only)	36	13 minutes
204a	Attitude toward Reading	1 form	9	untimed
204b	Attitude toward Mathematics	1 form	9	untimed
205a	Pupil Survey - Mathematics	1 form	11	untimed
205b	Pupil Survey - Reading	1 form	11	untimed
206	Reading Application			
	Fall	1 form	18	untimed
	Spring	2 forms	18	untimed
207	Decoding			
	Fall	5 forms		untimed
	Spring	4 forms		untimed
	Subtests:			
	Picture-Sound Correspondence		6	
	Rhymes		7	
	Word Recognition		7	
	Sight Words		14	
	Root Words		8	
	Words from Root Words		8	
	Word Sounds		60	
208	Mathematics Application	1 form	18	untimed
209	Reading Achievement			
	Fall	1 form	18	untimed
	Spring	2 forms	18	untimed
210	Short Form Test of Academic Aptitude			
	Fall only - Level 1			
	Subtests:			
	Vocabulary		19	untimed
	Analogies		20	7 minutes
	Sequence		20	untimed
	Memory		20	untimed
211	Embedded Figures Test			
	Spring only	1 form	25	untimed

*CAT raw scores are converted to normal curve equivalents before computing gains. See text for discussion.

Table 2

Pupil Test Battery - Grade 5

Test	Title	Form	Number of items	Timing
501	CAT Reading Comprehension			
	Fall - Level 3, Form A		36	35 minutes
	Spring - Level 3, Form B		36	35 minutes
502	CAT Mathematics Concepts			
	Fall - Level 3, Form A		25	7 minutes
	Spring - Level 3, Form B		25	7 minutes
503	CAT Mathematics Computation			
	Fall - Level 3, Form A		48	33 minutes
	Spring - Level 3, Form B		48	33 minutes
504a	Attitude toward Reading	1 form	14	untimed
504b	Attitude toward Mathematics	1 form	14	untimed
505a	Pupil Survey - Mathematics	1 form	6	untimed
505b	Pupil Survey - Reading	1 form	6	untimed
506	Reading Application			
	Fall	1 form	18	untimed
	Spring	2 forms	18	untimed
507	Decoding			
	Fall	5 forms		untimed
	Spring	4 forms		untimed
	Subtests:			
	Word Sounds 1		9	
	First Syllable		8	
	Last Syllable		8	
	Root Words		8	
	Words from Root Words		8	
	Word Sounds 2		60	
508	Mathematics Application	1 form	18	untimed
509	Reading Achievement			
	Fall	1 form		untimed
	Spring	2 forms		untimed
510	Short Form Test of Academic Aptitude			
	Fall only - Level 3			
	Subtests:			
	Vocabulary		25	7 minutes
	Analogies		20	7 minutes
	Sequences		20	12 minutes
	Memory		20	8 minutes
511	Embedded Figures Test			
	Spring only	1 form	25	untimed

Test Review, Try-Out and Revision

As the next step in the test development process, all instruments were reviewed by experts suggested by the Commission. These reviewers included:

Dr. Frederick Braun, University of Hawaii
Dr. Richard A. Denholm, University of California, Irvine
Dr. William Fisher, Superintendent, Orinda School District, California
Dr. Lloyd N. Morrisett, University of California, Los Angeles
Dr. Thomas A. Romberg, University of Wisconsin
Dr. Nicholas Silvaroli, Arizona State University

The comments of these experts, as well as the review and comments by Commission members and by ETS consultants, were of great help in revising the first draft of these tests.

Additionally, since it seemed important to find out how students would react to these test instruments, all of the newly developed measures were administered to one or more groups of children in the Princeton, New Jersey area.

Modifications of tests were made as a result of these two processes. The changes made involved better alignment of figures on the Elementary Group Embedded Figures Test, correction of typographical errors on the attitude tests, correction of typographical errors on the decoding tests and retyping of the grade 2 level tests on a primary typewriter, replacement of one selection from the grade 2 Reading Achievement Test because that selection discussed seasonal climate changes that might not be familiar to California children, modification of three selections in the grade 5 Reading Achievement Test and one in the grade 2 Reading Achievement Test primarily to shorten the selections and remove unusual words, replacement of three items on the grade 2 Reading Application Test that were too easy, replacement of one item on the grade 2 Reading Application

Test that was too difficult, modification of directions for one item on the grade 2 Reading Application Test, replacement of two selections on the grade 5 Reading Appliction Test that were too difficult, replacement of three items that were too easy on the grade 2 Mathematics Application Test, elimination of unnecessary reading from four items in the Mathematics Application Tests, replacement of two items in the grade 5 Mathematics Application Test that were too easy, and revision of one confusing item on the grade 5 Mathematics Application Test.

Following these changes the revised tests were administered to two classrooms of children in San Jose, California at a non-project school. These administrations were proctored by project staff members. Comments about the difficulty which some pupils had with the CAT answer sheets lead to a decision to administer all fall tests so that responses could be marked directly in the test booklets. The data from this try-out also suggested some further minor modifications in test content which were made before the final versions of the tests were printed.

A test administrator's manual was prepared for each grade level to facilitate uniform test administration.

Testing Procedures

Substitute teachers from each of the participating school districts were selected as testers. All testers were trained by ETS project staff members. Each tester supervised the two days of testing in a classroom. Records of testing conditions, unusual situations, problems encountered, and comments about the tests and

the testing conditions were kept for each classroom so that the effect of any irregularities on the test results would be known. Project staff serving as test monitors also reported on the testing situation and on the effectiveness of each tester.

The classroom schedule for test administration appears in Table 3. The schedule was planned to provide the pupils with adequate breaks and rest periods between tests and with enough variety from test to test so that boredom would not be a problem. The fall and spring test administration schedules were identical except that, in the spring, the Embedded Figures tests replaced the CAT Aptitude Tests. Some minor adjustments were made to fit school time schedules.

Additional Test Modification

After the fall testing was completed, all test results were examined so that any problems not detected by the review procedures could be corrected before the spring test administration. Few modifications were found to be necessary.

As was mentioned earlier, five alternate forms had been developed for the Decoding Tests administered in the fall. These forms were modified to create four forms for spring testing. The major change in making these forms was to eliminate certain words in the Word Sounds portion of this which were found to be poor distractors or to create confusion because of regional or dialect differences in pronunciation.

For the spring testing, as described previously, alternate forms of the Reading Application and Reading Comprehension Tests were developed to make it possible to determine the extent to which

Table 3

Schedule for Test Administration

Grade 2

Day 1	Day 2
CAT Reading Comprehension	Decoding Skills
(Recess)	(Recess)
CAT Math Concepts CAT Math Computation	Decoding Skills (if necessary) Math Application
(Recess)	(Recess)
Attitude and Pupil Surveys	Reading Achievement
(Lunch)	(Lunch)
Reading Application	CAT Short Form Aptitude (Fall) Fmbedded Figures (Spring)

Grade 5

Day 1	Day 2
CAT Reading Comprehension CAT Math Concepts	Decoding Skills
(Recess)	(Recess)
CAT Math Computation Attitude and Pupil Surveys	Math Application Reading Achievement
(Lunch)	(Lunch)
Reading Application	CAT Short Form Aptitude (Fall) Embedded Figures (Spring)

response differences between facets were item-specific. In so far as was possible, half of the items from the fall form appeared in each of the spring forms and old and new items from each test design facet were balanced in each of the alternate forms. In addition, some format changes were made in these spring forms of the tests, chiefly in the Reading Application items. The major effect of these format changes was to have the pupils mark one letter beside the correct option rather than to mark directly on the option itself.

Scoring

In the fall, pupils in both grade 2 and grade 5 marked their answers directly in the test books. This necessitated gridding of answers onto answer sheets before machine scoring. In the spring testing separate test answer sheets were used at the fifth grade level wherever possible to expedite the scoring. Because alternate forms of the California Achievement Tests were used for the fall and spring testing, it was necessary to equate the forms in order to compute growth in Reading Comprehension, Mathematics Concepts, and Mathematics Computation. It was also necessary to adjust the test score to take into account the differences in difficulty between the items administered and the items omitted in the scoring for this project. In the CAT Reading Comprehension Test, items on the use of a table of contents and index were not counted since they did not relate to comprehension as defined in this study. In grade 2, the multiplication and division items on the computation section were not counted. A correction, consisting of a constant which represents the assumed score on items not administered based on responses on the rest of the test, was added to the raw score of each test.

Specifically, the adjusted number right was computed as the number right plus the adjustment factor (for number and difficulty of items not taken). Then, using the CAT Achievement Development Scale Scoretable, the associated developmental score was found. Next, the percent right scores were obtained for each student and, finally, a gain score was computed. Details of the scoring procedure are given in Chapter 6 of Volume I.

Item Analysis

In analyzing the results of these tests, item anlyses were routinely performed on all newly developed or modified measures. The tests so analyzed included Reading Attitude, Mathematics Attitude, Reading Application, Decoding, Mathematics Application, Reading Achievement, and Embedded Figures. The item anlysis data included difficulty, point biserial, r biserial, and m coefficients for each item; a monotone transformation of the responses to each item and the monotone criterion for each option; and the mean total test score, standard deviation and percentage of students selecting each response option.

A summary of the item analysis data for the newly developed or modified test is presented in Table 4. It would be desirable to have dropped from future versions of these tests the few items which showed inadequate ability to discriminate (r biserials below .20) between high and low scorers. We recommended the replacement of one item from the grade 2 Reading Application Tests and three items from the grade 2 Reading Achievement Tests on the basis of the item analysis data. At grade 5, we recommended replacement of one item

on the Reading Attitude Test and the parallel item from the Mathematics Attitude Scale, two items from the Reading Application Tests, and five items from the Reading Achievement Tests. The report also notes that the two forms of the grade 5 Reading Achievement Test differ more in average difficulty (.48 for Form 1 and .53 for Form 2) than did the other alternate test forms developed in this study and that any replacement of the nondiscriminating items be done with attention to improving this imbalance.

Descriptive Statistics

The means, standard deviations, split-half reliabilities, and intercorrelations among the pupil test measures are presented in Tables 5 through 10. It should be noted that these means are raw scores and are not corrected for differences between forms; therefore, they should not be used to infer the amount of pupil growth.

The means and standard deviations indicate that the Reading Application Tests developed for this project were, for the most part, easier than the Reading Achievement Tests. There may have been some ceiling effect on the fall version of the fifth grade Reading Application Test and on the spring administration of the second grade Mathematics Application Test.

Table 4

Summary of Item Analysis Data

Grade	Measure	Range of Difficulties (p)		Mean Difficulty (p)	Range of r Biserials		Mean r Biserial
		Low	High		Low	High	
2	Reading Attitude	.48*	.82*	.67*	.39	.56	.46
2	Math Attitude	.45*	.84*	.67*	.37	.57	.46
5	Reading Attitude	.12*	.84*	.54*	.14	.54	.39
5	Math Attitude	.21*	.72*	.52*	.21	.58	.42
2	Reading Application (fall)	.28	.92	.65	.20	.69	.50
	- Form 1	.31	.87	.60	.18	.80	.60
	- Form 2	.30	.87	.60	.39	.79	.58
5	Reading Application (fall)	.28	.95	.75	.02	.56	.39
	- Form 1	.25	.85	.62	.14	.78	.50
	- Form 2	.33	.90	.59	.16	.72	.49
2	Decoding	.23	.98	.80	.22	.80	.54
5	Decoding	.38	.96	.78	.19	.89	.59
2	Math Application (fall)	.33	.98	.65	.12	.47	.29
	- (spring)	.45	.98	.74	.25	.53	.38
5	Math Application (fall)	.12	.91	.52	.30	.65	.45
	- (spring)	.12	.78	.52	.34	.67	.50
2	Reading Achievement (fall)	.23	.81	.52	.03	.61	.42
	- Form 1	.31	.85	.60	.08	.64	.45
	- Form 2	.33	.76	.58	.17	.73	.50
5	Reading Achievement (fall)	.33	.85	.51	.07	.60	.38
	- Form 1	.16	.83	.48	-.12	.57	.39
	- Form 2	.28	.73	.53	.15	.66	.43
2	Embedded Figures	.07	.93	.53	.49	.88	.65
5	Embedded Figures	.06	.83	.39	.46	.83	.66

* p choosing response keyed for positive attitude.

Table 5

Means ($\bar{X}$) and Standard Deviations (SD) for Grade 2
for Pupils with both Fall and Spring Scores

TEST		N	FALL 1973 $\bar{X}$	FALL 1973 SD	SPRING 1974 $\bar{X}$	SPRING 1974 SD
201	CAT Reading Comprehension	693	41.7	15.8	47.1	18.9
202	CAT Math Concepts	786	48.7	12.3	55.6	12.9
203	CAT Math Computations	769	38.1	12.4	46.5	16.5
204a	Attitude toward Reading	696	65.5	25.4	62.3	25.2
204b	Attitude toward Math	781	65.9	23.8	62.9	24.7
205a	Math Pupil Survey	784	72.9	19.5	83.2	14.5
205b	Reading Pupil Survey	697	74.6	17.4	79.2	15.2
206	Reading Applications	698	62.7	22.4	63.1	23.8
207	Decoding Total	698	70.3	15.2	78.9	13.4
208	Math Applications	785	65.5	15.2	75.6	14.6
209	Reading Achievement	694	51.7	22.8	62.4	23.8
210	Short Form Test of Academic Aptitude	797	59.2	18.6	--	--
211	Embedded Figures	797	--	--	54.4	23.9

Table 6

Reliability of Grade 2 Pupil Tests

Test	Title	Reliability Fall	Reliability Spring
201	CAT Reading Comprehension	.92	.94
202	CAT Math Concepts	.89	.64
203	CAT Math Computation	.95	.86
204a	Attitude toward Reading	.70	.69
204b	Attitude toward Math	.67	.68
204	Attitude Survey - Total	.81	.82
206a	Reading Application - Form 1	--	.84
206b	Reading Application - Form 2	--	.85
206	Reading Application	.84	--
207a	Decoding - Subtests a-g - Form 1	.96	.93
	Decoding - Subtests a-g - Form 2	.96	.93
	Decoding - Subtests a-g - Form 3	.95	.94
	Decoding - Subtests a-g - Form 4	.94	.92
	Decoding - Subtests a-g - Form 5	.95	--
207b	Decoding - Word Sounds - Form 1	.85	.95
	Decoding - Word Sounds - Form 2	.84	.95
	Decoding - Word Sounds - Form 3	.84	.95
	Decoding - Word Sounds - Form 4	.76	.94
	Decoding - Word Sounds - Form 5	.79	--
208	Mathematics Application	.63	.66
209a	Reading Achievement - Form 1	--	.77
209b	Reading Achievement - Form 2	--	.82
209	Reading Achievement	.79	--
210	SFIAA - Total	.94	--
211	Imbedded Figures	--	.90

Table 7

Intercorrelations for Grade 2 Test Battery
(Fall above the diagonal, spring below)

	201	202	203	204a	204b	205a	205b	206a	206b	206c	207a	207b	207c	207d	207e	207f	208	209a	209b	209c	210a	210b	210c	210d	210e	211
201 CAT Reading Comprehension	-	.68	.49	-.09	-.09	X	X	.72	X	X	.77	X	X	X	X	X	.50	.79	X	X	.54	.49	.35	.42	.30	X
202 CAT Math Concepts	.60	-	.60	-.03	-.07	X	X	.62	X	X	X	X	X	X	X	X	.63	.60	X	X	.70	.61	.46	.48	.33	X
203 CAT Math	.39	.43	-	-.01	-.04	X	X	.43	X	X	X	X	X	X	X	X	.51	.40	X	X	.52	.40	.38	.37	.21	X
204a Attitude Toward Reading	-.09	-.08	.01	-	.68	X	X	-.05	X	X	-.08	X	X	X	X	X	-.08	.11	X	X	-.06	-.11	-.06	.01	-.04	X
204b Attitude Toward Math	-.10	-.07	.03	.70	-	X	X	-.08	X	X	X	X	X	X	X	X	-.07	-.11	X	X	-.13	-.10	-.06	-.01	-.04	X
205a Pupil Survey-Math	.28	.36	.32	.05	.08	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
205b Pupil Survey-Reading	.41	.38	.25	.01	-.01	.48	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
206a Reading Application-Total	.69	.38	.33	-.10	-.12	.29	.40	-	X	X	X	X	X	X	X	X	.54	.67	X	X	.53	.48	.38	.40	.34	X
206b Reading Application-Form 1	.74	.60	.33	-.09	-.09	.35	.40	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
206c Reading Application-Form 2	.74	.62	.33	-.10	-.15	.23	.40	X	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
207a Decoding Subtests a-g	.63	.57	.36	-.08	-.01	.33	.36	.68	.73	.64	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
207b Decoding Word Sounds	.48	.38	.22	-.06	-.09	.15	.27	.48	.46	.49	.39	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X
207c Decoding Word Sounds-Form 1	.46	.37	.11	-.06	-.11	.08	.38	.57	.60	.56	.45	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X
207d Decoding Word Sounds-Form 2	.56	.45	.32	-.03	-.05	.25	.34	.51	.46	.56	.45	X	X	-	X	X	X	X	X	X	X	X	X	X	X	X
207e Decoding Word Sounds-Form 3	.51	.37	.28	-.11	-.13	.13	.29	.47	.47	.50	.36	X	X	X	-	X	X	X	X	X	X	X	X	X	X	X
207f Decoding Word Sounds-Form 4	.40	.34	.19	-.04	-.03	.16	.17	.40	.43	.37	.33	X	X	X	X	-	X	X	X	X	X	X	X	X	X	X
208 Math Application	.52	.58	.31	-.13	-.12	.28	.25	.55	.55	.55	.52	.32	.31	.34	.36	.28	-	.52	X	X	.60	.57	.45	.46	.37	X
209a Reading Achievement-Total	.67	.56	.26	-.08	-.09	.26	.38	.73	.75	.71	.58	.46	.51	.51	.48	.34	.51	-	X	X	.51	.49	.36	.44	.34	X
209b Reading Achievement-Form 1	.72	.56	.24	-.08	-.09	.24	.40	.73	.75	.72	.55	.47	.60	.48	.42	.38	.48	X	-	X	X	X	X	X	X	X
209c Reading Achievement-Form 2	.72	.57	.27	-.06	-.10	.28	.37	.73	.76	.71	.60	.45	.44	.54	.53	.31	.52	X	X	-	X	X	X	X	X	X
210a SFTAA-Total	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	.72	.74	.65	.49	X
210b SFTAA-Vocabulary	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	.43	.43	.25	X
210c SFTAA-Analogies	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	.50	.43	X
210d SFTAA-Sequence	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	.63	X
210e SFTAA-Memory	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	X
211 Embedded Figures	.43	.56	.43	-.05	-.02	.24	.23	.47	.49	.46	.45	.28	.31	.35	.22	.24	.49	.42	.37	.46	X	X	X	X	X	-

X = Not obtained or not completed.

Table 8

Means ($\bar{X}$) and Standard Deviations (SD) for Grade 5
for Pupils with both Fall and Spring Scores

TEST		N	FALL 1973 $\bar{X}$	FALL 1973 SD	SPRING 1974 $\bar{X}$	SPRING 1974 SD
301	CAT Reading Comprehension	1018	58.0	15.2	60.1	18.5
302	CAT Math Concepts	1121	55.2	13.4	56.6	14.8
303	CAT Math Computations	1128	43.8	12.6	46.3	14.5
304a	Attitude toward Reading	1014	57.4	22.3	55.9	23.5
304b	Attitude toward Math	1122	54.8	25.1	53.4	27.4
305a	Math Pupil Survey	1174	62.8	20.6	61.1	25.1
305b	Reading Pupil Survey	1058	70.1	15.7	64.9	22.5
306	Reading Applications	985	72.7	18.8	63.5	21.7
307	Decoding Total	1037	81.9	14.1	85.3	13.2
308	Math Applications	1151	55.2	20.2	52.6	21.8
309	Reading Achievement	1035	53.5	21.4	56.2	22.9
310	Short Form Test of Academic Aptitude	1273	47.4	17.9	--	--
311	Embedded Figures	1273	--	--	39.9	21.1

Table 9

Reliability of Grade 5 Pupil Tests

Test	Title	Reliability Fall	Reliability Spring
501	CAT Reading Comprehension	.93	.94
502	CAT Math Concepts	.84	.88
503	CAT Math Computation	.91	.90
504a	Attitude toward Reading	.75	.77
504b	Attitude toward Math	.77	.81
504	Attitude Survey - Total	.80	.82
506a	Reading Application - Form 1	--	.79
506b	Reading Application - Form 2	--	.78
506	Reading Application	.73	--
507a	Decoding - Subtests a-g - Form 1	.95	.92
	Decoding - Subtests a-g - Form 2	.94	.88
	Decoding - Subtests a-g - Form 3	.95	.90
	Decoding - Subtests a-g - Form 4	.95	.92
	Decoding - Subtests a-g - Form 5	.95	--
507b	Decoding - Word Sounds - Form 1	.89	.91
	Decoding - Word Sounds - Form 2	.89	.91
	Decoding - Word Sounds - Form 3	.91	.90
	Decoding - Word Sounds - Form 4	.87	.91
	Decoding - Word Sounds - Form 5	.88	--
508	Mathematics Application	.80	.82
509a	Reading Achievement - Form 1	--	.71
509b	Reading Achievement - Form 2	--	.76
509	Reading Achievement	.73	--
510	SFTAA - Total	.93	--
511	Embedded Figures	--	.88

Table 10

Intercorrelations for Grade 5 Test Battery
(Fall above the diagonal, spring below)

	501	502	503	504a	504b	505a	505b	506a	506b	506c	507a	507b	507c	507d	507e	507f	508	509a	509b	509c	510a	510b	510c	510d	510e	511
501 CAT Reading Comprehension	-	.70	.58	.23	.15	X	X	.60	X	X	.66	X	X	X	X	X	.65	.70	X	X	.80	.71	.46	.49	.56	X
502 CAT Math Concept	.70	-	.70	.27	.26	X	X	.61	X	X	X	X	X	X	X	X	.66	.56	X	X	.70	.43	.47	.55	.47	X
503 CAT Math	.62	.72	-	.28	.28	X	X	.52	X	X	X	X	X	X	X	X	.60	.51	X	X	.63	.49	.43	.56	.39	X
504a Attitude toward Reading	.33	.26	.16	-	.60	X	X	.20	X	X	.24	X	X	X	X	X	.20	.21	X	X	.25	.23	.25	.23	.18	X
504b Attitude toward Math	.17	.38	.36	.26	-	X	X	.11	X	X	X	X	X	X	X	X	.12	.14	X	X	.15	.14	.07	.15	.10	X
505a Pupil Survey-Math	.17	.24	.29	.16	.38	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
505b Pupil Survey-Reading	.19	.15	.05	.38	.16	.43	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
506a Reading Application-Total	.31	.66	.59	.30	.20	.17	.12	-	X	X	.60	X	X	X	X	X	.62	.56	X	X	.66	.34	.43	.51	.45	X
506b Reading Application-Form 1	.69	.67	.61	.31	.21	.16	.12	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
506c Reading Application-Form 2	.73	.67	.58	.29	.19	.18	.13	X	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
507a Decoding-Subtests a-g	.63	.67	.65	.32	.19	.18	.10	.63	.63	.64	-	X	X	X	X	X	X	X	.61	X	X	X	X	X	X	X
507b Decoding-Word Sounds Total	.57	.60	.53	.22	.16	.15	.10	.57	.55	.59	.74	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X
507c Decoding-Word Sounds-Form 1	.54	.57	.52	.20	.06	.11	.08	.51	.45	.58	.73	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X
507d Decoding- Word Sounds-Form 2	.57	.56	.50	.21	.16	.13	.15	.51	.51	.51	.71	X	X	-	X	X	X	X	X	X	X	X	X	X	X	X
507e Decoding-Word Sounds-Form 3	.58	.62	.50	.25	.22	.21	.08	.64	.65	.63	.74	X	X	X	-	X	X	X	X	X	X	X	X	X	X	X
507f Decoding-Word Sounds-Form 4	.60	.66	.58	.26	.19	.15	.09	.61	.59	.64	.78	X	X	X	X	-	X	X	X	X	X	X	X	X	X	X
508 Math Application	.73	.70	.64	.23	.31	.17	.10	.70	.69	.72	.64	.61	.59	.57	.64	.64	-	.60	X	X	.67	.60	.50	.53	.45	X
509a Reading Achievement-Total	.71	.60	.52	.31	.12	.12	.14	.63	.63	.64	.58	.50	.50	.47	.53	.51	.67	-	X	X	.67	.65	.46	.46	.53	X
509b Reading Achievement-Form 1	.72	.59	.49	.28	.12	.11	.10	.62	.59	.64	.52	.49	.51	.47	.48	.49	.66	X	-	X	X	X	X	X	X	X
509c Reading Achievement-Form 2	.74	.61	.56	.33	.12	.12	.16	.66	.67	.65	.53	.53	.51	.47	.58	.55	.68	X	X	-	X	X	X	X	X	X
510a SFTAA - Total	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	.79	.77	.81	.60	X
510b SFTAA - Vocabulary	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	.43	.46	.37	X
510c SFTAA - Analogies	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	.57	.47	X
510d SFTAA - Sequence	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	.55	X
510e SFTAA -	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	X
511 Embedded Figures	.48	.52	.44	.06	.16	.12	.03	.44	.47	.47	.47	.46	.44	.43	.48	.49	.57	.44	.44	.45	X	X	X	XX	X	-

X = not obtained or not computed.

As can be seen from Tables 6 and 9, the reliabilities of the newly developed pupil measures ranged from .63 to .95; the reliabilities of the CAT tests ranged from .64 to .95. Therefore, the new tests can be considered comparable in reliabilities to these formally published standardized tests. Given the fact that many of the newly developed tests were fairly short and that their reliabilities could be boosted by increasing their length (such as by combining items from the alternate forms), the tests seem adequately reliable.

The attitude scales have reliabilities ranging from .67 to .82; the lower reliabilities are associated with the shorter (18 item) grade 2 scales while the higher reliabilities are associated with the slightly longer (28 item) grade 5 scales.

The 18 item Reading Application Tests had reliabilities ranging from .73 to .85 while the 18 item Reading Achievement Tests had reliabilities ranging from .71 to .82. Both of these tests have alternate forms at each of the two grade levels which could be combined to produce longer, more reliable measures.

The reliabilities of the Decoding Tests range from .76 to .95. These tests are already fairly long and had, by the time the spring revised versions were administered, reliabilities over .90.

The Mathematics Application Tests, which were also 18 items long, had reliabilities of .63 and .66 at the grade 2 level and .80 and .82 at grade 5. It was in grade 2 that the CAT Mathematical Concepts Test also gave the surprisingly low reliability of .64. Perhaps the second graders in this project really were erratic in their handling of mathematics!

The intercorrelations shown in Tables 7 and 10 indicate that, despite the test construction efforts to make distinctively different reading tests, the reading measures are highly intercorrelated. This suggests that reading comprehension skills may not be multi-dimensional, at least not in the ways conceptualized in this study. At the second grade level especially, attitudes toward reading and mathematics are highly correlated suggesting that attitude toward school may be fairly global at this age and becomes more differentiated with increasing school experience.

Relationship Between the Tests and Instruction

There has been considerable interest in the extent to which tests used in studies of this sort are really sensitive to the opportunity to learn and to teachers' instructional emphases. Consequently, the correlations between the gains (fall to spring) on the pupil tests and the time devoted to and the frequency of teaching certain reading and mathematics skills (as obtained from the Work Diaries) were computed. These are shown in Tables 11 and 12. In general, the time and frequency of teaching a skill do not seem to be highly correlated with related pupil test gains.

While the most highly significant correlations do not show a perfect correspondence between teaching emphasis in reading and mathematics and pupil gain, there is a general tendency for the frequency with which a skill is taught to be positively related to pupil gains. The amount of time a skill is taught appears to have less effect on pupil test gains suggesting, once more, the importance of distributed practice. A further discussion of related data can be found in the report on the Work Diary, Volume V.5.

Table 11

Correlations Between Amount of Time and Frequency of Teaching Certain Reading Skills and Pupil Test Mean Gain

Grade	Skill Taught		CAT Reading Comprehension	Reading Application	Pupil Tests Decoding	Reading Achievement	Total
2 (n = 43)	Decoding	- Time	-.01	.29*	.18	.08	.22
		- Frequency	.12	.18	.17	.27*	.07
	Vocabulary	- Time	-.00	-.09	.22	.04	-.07
		- Frequency	.05	.13	.06	.19	-.19
	Comprehension	- Time	-.13	-.06	.12	.13	-.12
		- Frequency	.21	.26*	.18	.10	.38**
	Application	- Time	.03	-.00	-.14	.03	.06
		- Frequency	.36**	.03	.15	.32**	.20
5 (n = 5)	Decoding	- Time	-.42***	-.19	.05	-.24*	-.39**
		- Frequency	-.08	-.00	.17	.06	-.17
	Vocabulary	- Time	-.35**	.10	.17	-.25*	-.18
		- Frequency	-.08	.20	.09	.22	-.20
	Comprehension	- Time	.01	-.18	-.23*	-.15	-.19
		- Frequency	.03	.01	-.11	-.02	.14
	Application	- Time	-.10	-.09	-.21	-.06	-.16
		- Frequency	.12	.02	.01	.01	-.04

*p < .10 **p < .05 ***p < .01

Table 12

Correlations Between Amount of Time and Frequency of Teaching Certain Mathematics Skills and Pupil Test Gains

Skill Taught	Pupil Tests: CAT Math Concepts	CAT Math Computation	Math Applications	Math Total
Grade 2				
Problems - Time	-.08	-.07	.21	.04
- Frequency	.27	.02	.25	.24
Operations - Time	.03	-.03	.08	.01
- Frequency	-.03	.44	.17	.31
Special Skills - Time	.33	-.18	.13	.04
- Frequency	.11	.04	.08	.13
Grade 5				
Problems - Time	-.21	-.04	-.02	-.10
- Frequency	-.14	-.03	.15	.04
Operations - Time	-.13	.10	.01	-.02
- Frequency	-.15	-.05	.15	.03
Special Skills - Time	-.06	.03	.04	.01
- Frequency	-.38	-.19	.11	-.08

Teacher Measures

In this section the rationale behind teacher measures will be described, the process of selecting and developing the tests outlined, and the statistical analysis procedures and descriptive statistics relating to these instruments presented.

There are a variety of teacher measures which have been included in this study. Among them are tests of teacher aptitude and cognitive style, teaching and subject matter knowledge, and measures of teacher attitudes toward teaching.

Relevance of Teacher Aptitude

Recognition of the importance of teacher aptitude as a part of this study can be traced back to the research planning conference held by the California Commission for Teacher Preparation and Licensing in February 1973. As the report of the pre-conference planning indicated, one of the objectives of the research being proposed was "to explore thoroughly verbal ability as a positive teacher variable in order to isolate the competencies involved which cause verbal ability to emerge as a consistent predictor of pupil achievement."

The Research Design Report, prepared for the Commission by National Education Program Associates (NEPA), pointed out the importance of measuring both teacher verbal ability and teacher quantitative ability. As the review of the literature, also prepared by NEPA, pointed out, "teacher verbal ability has been found to be related to student achievement within the Coleman (1966) study." Several other studies have confirmed that teacher verbal ability

is related to student achievement (Bowles, 1968; Bowles and Levin, 1968; Hanushek, 1968; Hanushek, 1970). The NEPA report also described a review of the research by Morsh (1954) which suggested that there was a low positive correlation between teacher intelligence and success in teaching. The degree to which aptitudes, as sub-components of the concept of general intelligence, might be differentially related to teaching success in certain subject matter areas does not seem to have been explored in the NEPA materials.

However, the teacher measures suggested in the Research Design Report gave an indication that a broader conceptualization of teacher aptitude might be appropriate. For example, the Cornell Critical Thinking Test was suggested as a measure of verbal fluency and the Cognitive Abilities Tests as a measure of quantitative aptitude. Consequently, in the ETS proposal, it was suggested that a broader array of teacher aptitudes be examined in this study.

The rationale behind introducing a broader consideration of teacher aptitudes in the ETS proposal was that: (1) although research relating one aptitude, teacher verbal ability, to pupil achievement had shown a fairly strong relationship, little or no research had been done to determine if other teacher aptitudes might also affect pupil learning, and (2) it seemed likely that there would be an interaction between both teacher knowledge and performance and teacher aptitudes. The basic relationship between aptitude and achievement has been quite thoroughly explored in much other research. Consequently, our hypotheses that both teacher knowledge and teacher performance are at least partially dependent upon teacher aptitude can hardly be considered revolutionary.

Selecting Teacher Aptitude Dimensions

As a first step in selecting teacher aptitudes for inclusion in this study, the literature on the relationship between teacher aptitudes and pupil performance was reviewed.

Verbal ability is the only teacher aptitude which has received much research attention. Most of this interest stems from the finding of the Coleman study (Equality of Educational Opportunity, 1966) which showed that teachers' verbal ability was one of the characteristics most consistently related to pupil achievement. This finding has been replicated by other researchers (e.g., Hanushek, 1970).

We do not know, however, why teacher verbal ability is important to pupil achievement. A number of reasons have been suggested. These include hypotheses ranging from suggestions that verbal ability is simply a proxy for general intelligence to suggestions that teachers with higher verbal ability are more able to negotiate their way into schools where pupils overachieve for other reasons. It seems likely that the size of a teacher's vocabulary may be related to her/his success in communicating with students. However, this relationship may not be the same for all groups of students. For example, in working with Mexican-American children with a more limited knowledge of English, the larger teacher vocabulary may be a source of confusion rather than providing multiple avenues to understanding.

An interesting related problem is whether the relationship between teacher verbal ability and pupil achievement is linear or curvilinear.

Taylor et al (1967) has suggested that people with high verbal ability are able to communicate more ideas in a given perio' of time but it is possible to think of situations where the pacing of new ideas could be too rapid for some groups of pupils, especially if accompanied by unfamiliar vocabulary.

Other literature suggested that skills in the divergent production aptitudes which are related to verbal aptitude, such as fluency and flexibility, might also be important in teaching.

Both associational and expressional fluency have been shown to be skills utilized when instructing others and, therefore, directly related to pupil achievement. However, individuals with very high or very low ideational fluency have been found to be less effective in transmitting information than those with ideational fluency in the middle range (Taylor et al, 1967).

Student teachers who are more flexible have been shown to be more effective in helping children to think for themselves (Hunt and Joyce, 1967). Good communicators are more flexible than poor communicators, but flexible individuals may be judged by others as performing in ways that are less efficient and less effective (Taylor et al, 1967). It has also been suggested that there may be a systematic interaction between convergent and divergent teachers and convergent and divergent pupils (Hudson, 1968).

Another source of information which served as a basis for the selection of teacher aptitudes to be included in this research were various models of the structure of intellect (Carroll, 1974; Guilford, 1967) and other research and reviews of cognitive factors (e.g. Ekstrom, 1973). After considering all of these, cognitive factors which

appeared to have relevance for the teaching skills necessary in elementary school reading and mathematics were selected for investigation. These factors were: verbal aptitude; quantitative aptitude; inductive and logical reasoning; associative and rote memory; associational, expressional and ideational fluency; semantic originality; semantic redefinition; sensitivity to problems; and spontaneous semantic flexibility.

Tests Selected for Aptitude Measurement

Most of the tests selected for use with the teachers in this study were adaptations of instruments from the ETS Kit of Reference Tests for Cognitive Factors (1963) or from similar measures being developed for the 1976 revision of this kit. The reason for this was that these tests are designed to be essentially univocal measures of a single aptitude.

Two tests not part of this kit of tests were also included in the teacher aptitude battery. One was the meaningful memory test (Sentence Completion Test) which had been developed in other research at ETS. The other was the verbal ability test which had been developed by ETS for the Coleman study. It seemed particularly important to include this instrument to serve as a "marker" measure so that the findings of this study could be related to other research using this same test as an index of teacher verbal ability. As the later analyses will show, its inclusion in this more complex battery has provided data which suggests that this instrument is much more nearly a measure of verbal reasoning than a simple vocabulary test. Perhaps, this explains why some researchers have found teacher scores on the Coleman study test

to be more highly correlated with pupil achievement in mathematics than with achievement in reading. A description of the aptitude measures, and the other teacher test scores, is given in Table 13.

Teacher Knowledge

In developing this model, two major components of teacher knowledge were hypothesized: (1) knowledge of the subject being taught and (2) knowledge of teaching.

Knowledge of subject matter is a familiar concept; however, the relationship between teacher knowledge in a content area and pupil achievement in that same area may not necessarily be a positive one. For example, one study (Rouse, 1968) found a slightly negative correlation between the amount of college mathematics studied by elementary school teachers and the arithmetic achievement of their pupils.

In this study, we defined knowledge of reading and of arithmetic in terms of the teacher's ability to do the kinds of decoding, inferential reading and arithmetic computation required of their pupils.

The knowledge of teaching is a somewhat more complex concept than subject knowledge and involves not only the techniques of classroom management but also the methods and materials which can be used in teaching a given subject. This concept also includes the ability to know what methods and materials should be applied to different students and when each is appropriate or inappropriate.

In the BTES study, we attempted to measure both of these components of teacher knowledge. Since this research is focused

on the teaching of reading and mathematics, we were concerned about the teachers' own skills in these subjects. Consequently, we tested teachers' mathematics knowledge by asking them to do both simple arithmetic computations and also to solve more complex mathematical problems. While it would have been desirable, if testing time had allowed, to measure these teachers' knowledge of more advanced mathematics, the rationale was to measure the arithmetic knowledge and skills which these teachers had to employ on a day-to-day basis in instructing their pupils as being the most directly relevant to the teaching process. We measured reading skill by giving each teacher a test of decoding skills and in addition, a measure to assess a higher level comprehension skill, drawing inferences from reading. We assumed that a teacher who had trouble with determining letter-sound correspondence or doing simple computations might be less effective in teaching that skill than a teacher with more skill in these areas. This could be because the teacher is less able to communicate the essential knowledge to the pupil or, because the teacher spends less instructional time on areas in which she/he feels less competent or less comfortable, thereby providing the pupils less opportunity to learn.

We also administered tests of the teachers' knowledge of the teaching of reading and mathematics. The reading tests also included teachers' application of their knowledge of teaching reading to certain hypothetical situations. The knowledge of teaching mathematics test contains some questions which are indicative of the extent of the mathematics background of the teacher as well as questions of teaching methodologies relevant to mathematics.

Table 13

Teacher Test Scores

APTITUDE SCORES:

Verbal Fluency - A composite score, derived from a factor having its major loadings on:

701	Verbal Facility	--	The verbal ability test used in the Coleman study. The subject is asked to select, from among five options, the best word to be used to complete a sentence.
702	Vocabulary	V-4	The subject is asked to select, from five options, the best synonym for a stimulus word.
725	Topics	FI-1	An ideational fluency test. The subject is asked to write as many ideas as possible about a given topic.
727	Controlled Associations	FA-1	An associational fluency test. The subject is asked to write as many synonyms as possible for each of several stimulus words.
728	Making Sentences	FE-1	An expressional fluency test. The subject is asked to write sentences of a designated length when the first letter of most of the words is specified.

Memory - A score derived from a doublet factor with loadings on:

712	Picture-Number	MA-1	An associative memory test which asks the subject to recall the two-digit numbers paired with each of several pictures previously studied.
724	Sentence Completion	--	A meaningful memory test in which the subject is asked to recall the one word which has been deleted from each of several previously studied sentences.

Reasoning - A composite score derived from a factor having its major loadings on:

713	Picture Grouping	I-3	A figure classification test. The subject is asked to determine the rule or reason which determines the assignment of a simple figure to one of two or three groups and then to indicate to which group additional figures should be assigned.
714	Nonsense Sayings	RL-1	A logical reasoning test. The subject is asked to indicate if the conclusion drawn from two preceding statements shows good or poor reasoning.
715	Inductive Reasoning	--	The subject is asked to determine the rule which relates four groups of four letters and to mark the fifth and unrelated group.
717	Mathematics Aptitude	RG-2	The subject is asked to select from five options the correct answer to simple word problems which stress reasoning and include some simple algebra.

Flexibility - A composite score based on a factor with its main loadings on:				
	721	Finding Useful Parts	RE-1	A test of redefinition ability. The subject is asked to select, from five options, the one object which could be used as a make-shift substitute for a specified purpose when the object usually used is unavailable.
	726	Story Surprises	O-1	A test of semantic originality. The subject is asked to write two different and surprising endings for each of several short stories.
	729	Listing Objects	XS-3	A test of semantic flexibility. The subject is asked to list as many things as possible which might be found in a specified setting.
	730	Planning Test	Sep-1	A test of sensitivity to problems. The subject is asked to indicate what is wrong with each of several plans presented for solving a variety of practical problems.
COGNITIVE STYLE:				
	723	Score on the Group Embedded Figures Test		
KNOWLEDGE SCORES:				
		Teaching Methods		Scores derived from short tests of methods of teaching reading or mathematics at the elementary school level. Some items include general theoretical background in the subject.
		Subject Matter		Scores derived from teacher performance on tests of inferential reading and decoding or of simple arithmetic computation.
ATTITUDE SCORES:				
		Aspirations		A composite score derived from six items relating to desire for leadership, recognition, and opportunities.
		Satisfaction		A composite score derived from ten items dealing with satisfaction with various aspects of the school, with teaching as an occupation, and with contacts with teachers and administration.
		Perception of Student Characteristics		A composite score derived from three items about student educational background, socioeconomic level, and difficulty in controlling students.

Implications

In looking at both teacher knowledge and teacher performance, we were concerned with the basic aptitudes and skills required. We hypothesized that reasoning as well as verbal aptitudes would be an important component in knowledge of the subject matter areas being studied. For example, in reading comprehension, inferential reasoning is one of these basic aptitudes. It seems possible that a teacher whose own inferential aptitude is weak would be less effective at developing interential skills in pupils than a teacher with more aptitude in this area.

The most likely model for the relationship between teacher aptitudes, knowledge, and performances is that there exists some minimal aptitude level which is necessary but not sufficient for the acquistion of knowledge, both about the subject matter area and about teaching skills, and that this knowledge will produce adequate pupil acquisition of information and adequate teaching performance. The aptitudes specific to certain subjects are probably moderated by the teacher's communication techniques which are also probably partially dependent on other aptitudes, such as verbal facility.

There are practical as well as theoretical reasons why teacher aptitudes were considered to be an important part of this study. Since the ultimate outcome of this research is to suggest possible modifications in teacher licensing standards, it seems desirable to have variables which can be readily quantified for use as selection devices. The importance of focusing on teacher abilities, as opposed to variables such as teaching style or teacher enthusiasm, was

pointed out by several of the consultants attending the planning conference as being particularly appropriate to the long-range goals of this research.

This approach also follows the model, long used in industrial psychology, which first requires an analysis of the task or job being studied and then the relationship of the task elements to known aptitudes and abilities. This approach allows the prospective employer both to select individuals on the basis of existing skills and also to select for training those individuals who have the greatest likelihood of being successful in programs which will train them to develop those skills.

Cognitive Style

Cognitive style has been shown to be an important variable in school performance (Witkin, 1973). It not only affects pupils' academic development at every educational level, but also is an important factor in understanding how students learn, how teachers teach, and the classroom interaction of teachers and students.

We know that people who choose elementary school teaching as a profession tend to be relatively field-dependent. Preferred teaching strategy is also related to cognitive style. The more field-dependent teachers tend to use discovery or discussion methods while field-independent teachers tend to prefer lecturing (Wu, 1967). But perhaps the most interesting finding has been that, when teachers and pupils are matched for cognitive style, they describe each other more positively than do mismatched teachers and pupils (Di Stefano, 1969). Witkin (1973) has hypothesized that people with matched cognitive styles

may get along together better for three reasons: (1) shared foci of interest, (2) common person characteristics, and (3) similarity of communications modes. The question remains as to whether this enhanced interpersonal perception has any consequences in terms of pupil achievement. Research (Thelen, 1967) which tried to group classes according to teachers' perception of the "teachability" of the students, showed that while students with high "teachability" scores got higher grades than low "teachability" students, there were no differences in achievement test gains.

Pilot Testing and Revisions

The draft versions of aptitude and knowledge tests were tried out on a small sample of Rhode Island elementary school teachers. Minor changes, primarily involving format and directions, were made and the tests were then submitted to the Commission's Evaluation of Teaching Committee and to their Research Advisory Board for review. The membership of these groups was not identical to those reviewing the pupil measures but there was some overlap. Final revisions were made as a result of the comments received from these reviews.

Testing Procedures

Because it was assumed that teacher aptitudes, knowledge, and attitudes toward reading and mathematics were relatively stable, each of these was measured only once. The teacher tests were administered to the teachers in small groups by project staff on the days when their classes were taking the pupil test battery. The fall tests battery consisted of three booklets:

I. Teaching of Reading

A. Verbal aptitude

1. Verbal facility test from the Coleman study

2. Vocabulary test developed by ETS

B. Decoding Skills - tests developed by Robert Calfee

1. Letter-sound correspondence

2. Root words

3. Affixes

4. Syllabification

C. Inferential Reading - test developed by ETS

D. Knowledge of Teaching Reading - tests developed by Robert Calfee

1. Primary knowledge

2. Primary applications

3. Intermediate knowledge

4. Intermediate application

II. Organization of Information (Aptitudes) - tests developed by ETS

A. Memory

B. Figure classification

C. Deductive reasoning

D. Inductive reasoning

III. Teaching of Mathematics - tests developed by ETS

A. Computations

B. Mathematics aptitude

C. Knowledge of teaching mathematics

The spring teacher test battery consisted primarily of tests of various divergent production abilities developed by ETS. The teacher cognitive style measure (Group Embedded Figures Test, published by Consulting Psychologists Press) was administered in the spring. As mentioned previously, all of the teacher measures are described in Table 13.

Scoring and Data Reduction

Because of the small size of the teacher sample (N=95), these measures were hand scored.

In order to reduce the number of teacher aptitude variables to a manageable number for inclusion in the path analysis, the data from these tests were factor analyzed. The resulting teacher aptitude factors were:

I - Verbal Fluency, defined as the ability to use words skillfully. The tests with the major loadings on this factor are measures of vocabulary and fluency.

II - Memory, a doublet factor including memory for both verbal and nonverbal material.

III - Reasoning, defined as the ability to solve problems requiring inductive, syllogistic, and mathematical reasoning.

IV - Flexibility, defined as the ability to change set. The tests with major loadings on this factor all require breaking one set and producing another solution.

A more complete description of the factor analysis may be found in the teacher aptitude substudy (see Volume V.1.). The composition of these factors is also indicated in Table 13.

Descriptive Statistics

The data on the teacher tests administered in the fall is presented in Table 14. As can be seen, the reliabilities of these measures varied considerably with a range from -.19 to .96. However, only two tests had very low reliabilities: (1) the Affixes Test (.17) and (2) Primary Reading Knowledge Test (-.19). The reason for the low reliability on both measures is undoubtedly related to the very limited amount of variance on these measures. Three other teacher tests had reliabilities in the .40's. These were the Root Words Test, the Intermediate Reading Knowledge Test, and the Mathematics Teaching Knowledge Test. The tests of knowledge of subject area (decoding, inferential reading, mathematics computation) had reliabilities ranging from .59 to .96. The teacher aptitude measures had reliabilites ranging from .67 to .94. It should be noted that many of these fall teacher tests are quite short and, thus, an increase in the length of these measures could be expected to improve their reliability.

The means and standard deviations for the teacher tests administered in the spring are given in Table 15. Reliabilities are not available for these measures because only total scores were obtained and not item data. Traditional items analysis techniques are not appropriate to divergent production meausres. However, communalities from the

Table 14

Test Statistics for Fall Teacher Measures

Test Title	Number of Teachers	Number of Items	Total Group Scores-Mean	Total Group Scores-S.D.	Reliability
VERBAL FACILITY	97	30	24.94	3.36	.72
VOCABULARY	97	18	11.64	3.05	.77
DECODING	97	300	281.84	17.95	**
ROOT WORDS	97	45	44.87	0.53	.54
AFFIXES	97	15	14.97	0.17	**
SYLLABIFICATION	97	30	27.24	1.69	.50
INFERENCE	97	10	6.83	2.00	.59
PRIMARY KNOWLEDGE	42	6	4.57	0.79	*
PRIMARY APPLICATION	42	35	14.88	3.23	.53
INTERMEDIATE KNOWLEDGE	55	7	5.00	1.28	*
INTERMEDIATE APPLICATION	55	35	13.62	3.42	.67
MEMORY	97	21	13.06	4.86	.84
PICTURE GROUPING	97	112	52.92	15.80	.48***
NONSENSE SAYINGS	97	15	6.88	2.77	.67
INDUCTIVE REASONING	97	15	10.48	2.57	.71
MATHEMATICS COMPUTATION	97	80	32.64	7.90	.79***
MATHEMATICS APTITUDE	97	21	7.75	3.55	.74
TEACHING MATH	97	16	8.89	2.49	.51

* = Reliabilities not computed due to short length of these tests.
** = Reliability not computed due to lack of variance on this measure.
*** = Communalities which represent lower boundary reliability estimates.

Table 15

Test Statistics for Spring Teacher Measures

Test Title	# of Teachers	# of Items	Mean	Standard Deviation	Communalities* Grade 2	Communalities* Grade 5
Finding Useful Parts	94	10	6.98	2.07	.70	.63
Embedded Figures	94	25	17.98	4.95	.66	.71
Sentence Completion	94	70	26.90	12.00	.75	.72
Topics	94	**	27.52	6.19	.67	.76
Story Surprises	94	6	3.49	2.16	.70	.64
Controlled Associations	90	8	41.86	13.60	.67	.54
Making Sentences	94	14	11.03	2.59	.86	.69
Listing Objects	94	**	25.60	5.50	.74	.68
Planning	94	**	7.28	1.92	.84	.64
Attitude - Teaching Reading	94	10	8.67	1.54	.82	.82
Attitude - Teaching Mathematics	94	10	8.96	1.81	.74	.77

* Communalities represent a lower-boundary reliability estimate.
** Free response items in test; maximum score not known.

teacher aptitude test factor analysis provided a lower bound estimate of reliability for most of these measures. As can be seen, these reliability estimates range from a low of .49 to a high of .78. Again, these measures are fairly short in length and could be made longer to improve their reliability.

As these tables also show, there is a ceiling effect on several of the teacher knowledge tests, especially for the measures related to knowledge of teaching and to decoding skills.

The intercorrelations among the teacher knowledge, aptitude, and attitude measures are shown in Table 16. These are based on data for the total group of teachers. For a discussion of differences in the relationships among the aptitude variables between grade 2 and grade 5 teachers, the reader is referred to the teacher aptitude substudy in Volume V.1.

Table 16

Teacher Test Intercorrelations

		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	21	23	24	25	26	27	28	29	30	31-1	31-2
701	Verbal Facility	--	.54		.19	.10	.17	.69	.16		.43		.24	.26	.22	.89	.18	.43	.41	.54	.45	.42	.22	.47	.35	.29	.16	.43	-.14	-.15
702	Vocabulary	.54	--		.23	.08	.13	.58	.16		.43		.16	.22	.01	.31	.13	.32	.35	.36	.37	.16	.31	.29	.50	.35	.14	.25	-.10	-.04
703	Decoding																													
704	Root Words	.19	.23		--	-.04	.13	.28	.16		.23		.09	.19	-.03	.38	.02	.20	.19	.31	.27	.26	.08	-.04	.17	.14	.18	.16	-.10	-.12
705	Affixes*	.10	.08		-.04	--	.02	.10	.11		.00		.03	-.10	-.12	.06	.22	-.03	.02	.15	.01	-.15	-.07	-.04	.12	-.02	-.06	.03	.04	.06
706	Syllabification	.17	.13		.13	.02	--	.12	-.35		.21		-.16	.07	.03	.15	.04	.06	.21	.11	.12	.03	-.18	.10	.08	.10	.06	.06	-.17	-.23
707	Inference	.69	.58		.28	.10	.18	--	.24		.45		.26	.42	.17	.54	.23	.47	.52	.47	.54	.39	.22	.46	.45	.45	.20	.57	-.18	-.10
708	Primary Knowledge**	.16	.16		.16	.11	-.35	.24	--		.00		.27	.08	-.06	.11	-.03	.09	-.09	.03	.06	.06	.29	-.02	.01	-.15	-.09	.29	-.16	.06
709	Primary Application																													
710	Intermediate Knowledge**	.43	.43		.28	.00	.21	.45	.00		--		.14	.29	-.05	.37	-.14	.19	.51	.42	.53	.07	.13	.11	.36	.23	.20	.37	-.08	-.17
711	Intermediate Application																													
712	Memory	.24	.16		.09	.03	-.16	.26	.22		.14		--	.23	.03	.28	.10	.30	.20	.20	.29	.50	.04	.24	.04	.11	.18	.13	-.16	-.20
713	Picture Grouping	.28	.22		.19	-.10	.07	.42	.07		.29		.23	--	.16	.46	.21	.33	.27	.27	.46	.21	.12	.31	.15	.16	.12	.20	-.19	-.16
714	Nonsense Sayings	.23	.01		-.02	-.11	.03	.17	-.06		-.05		.03	.16	--	.19	.23	.31	.24	.08	.02	.19	.12	.18	.13	.11	.13	.13	-.17	-.12
715	Inductive Reasoning	.39	.31		.38	.06	.15	.54	.11		.37		.28	.48	.19	--	.21	.45	.47	.42	.56	.30	.14	.42	.30	.40	.30	.42	-.16	-.11
716	Math Computation	.18	.18		.02	.22	.04	.23	-.03		-.14		.10	.21	.23	.21	--	.35	.14	.10	.21	.03	.24	.09	.33	.23	.24	.09	-.18	-.16
717	Math Application	.43	.32		.20	-.03	.06	.47	.09		.19		.30	.37	.31	.45	.35	--	.48	.30	.51	.20	.31	.34	.38	.26	.19	.32	-.21	-.16
718	Teaching Math	.41	.35		.19	.02	.21	.56	-.09		.51		.20	.27	.24	.47	.14	.48	--	.29	.46	.26	.11	.46	.44	.35	.24	.47	-.18	-.11
721	Useful Parts	.54	.36		.31	.19	.10	.47	.03		.42		.26	.27	.08	.42	.20	.30	.29	--	.41	.35	.15	.33	.19	.25	.07	.35	-.27	-.29
723	Embedded Figures	.45	.37		.27	.01	.12	.54	.06		.53		.29	.40	.02	.58	.21	.51	.46	.41	--	.31	.18	.41	.45	.27	.16	.33	-.28	-.26
724	Sentence Completion	.42	.16		.26	-.15	.03	.39	.06		.07		.50	.29	.19	.30	.03	.20	.14	.35	.31	--	-.03	.27	.11	.11	.24	.22	-.26	-.19
725	Topics (I)	.22	.31		.08	-.03	-.18	.22	.29		.13		.09	.12	.12	.14	.24	.31	.11	.15	.18	-.09	--	.16	.33	.34	.22	.62	-.04	-.04
726	Story Surprises	.47	.29		-.04	-.04	.10	.46	-.02		.11		.29	.31	.18	.42	.09	.34	.46	.33	.41	.27	.16	--	.26	.43	.21	.35	-.25	-.20
727	Controlled Associations	.35	.50		.17	.12	.03	.45	.01		.36		.04	.15	.13	.30	.33	.38	.44	.18	.45	.11	.33	.26	--	.36	.34	.20	-.16	-.12
728	Making Sentences	.24	.35		.14	-.02	.10	.45	-.15		.13		.11	.16	.11	.40	.25	.26	.35	.25	.27	.11	.34	.43	.36	--	.23	.40	-.25	-.26
729	Listing Objects	.16	.14		.18	-.08	.08	.20	-.04		.10		.18	.22	.13	.36	.28	.16	.20	.07	.26	.29	.21	.21	.34	.23	--	-.02	-.20	-.21
730	Planning Test	.43	.28		.19	.03	.06	.57	.24		.37		.13	.20	.13	.41	.04	.33	.46	.38	.33	.26	.22	.36	.20	.40	-.07	--	-.09	-.07
731-1	Attitude - Reading	.19	-.16		-.10	.24	-.17	-.18	-.08		-.06		-.26	-.24	-.17	-.28	-.18	-.21	-.18	-.27	-.28	-.26	-.04	-.15	-.06	-.15	-.10	-.04	--	.28
731-2	Attitude - Mathematics	.15	-.04		-.12	.08	-.23	-.16	.04		-.13		-.09	-.04	-.11	-.11	.07	.02	.01	-.09	-.05	-.23	.04	-.09	-.15	-.01	-.11	-.07	.28	--

* = Correlations for this test should be interpreted with caution due to the low reliability of the test.

** = Correlations for this test should be interpreted with caution due to the few items involved.

Other Measures

In addition to the pupil and teacher measures discussed in the two preceding sections, several other types of measures were included in this study. They were: (1) a diagnostic film test, (2) a teacher background questionnaire, (3) a questionnaire for teachers on the school setting, (4) an instrument developed by Rensis Likert to elicit teacher and principal perceptions of school climate, (5) an instrument designed to elicit teachers' expectations for pupil performance, (6) a pupil survey questionnaire completed by the teacher, (7) a form with information about pupil's native language and attendance completed by the teacher, (8) a parent questionnaire for pupil and family background information, (9) a questionnaire for principals on school characteristics, and (10) a structured interview with the principals.

Diagnostic Film Test

Because it seemed important to include a more realistic measure of teaching knowledge than could be obtained with any paper-and-pencil test, a film test was designed to assess teachers' ability to diagnose pupil reading difficulties. This test is a series of videotaped episodes depicting children reading either from word lists or paragraphs. Each child demonstrates some difficulty in performing a particular reading skill. The teachers were asked to diagnose the reading problem shown and to recommend strategies for its correction. The test was administered to 17 second grade teachers and 17 fifth grade teachers. Because not all the teachers took this measure, it

was not included in subsequent analyses. A more complete description of this measures and the descriptive statistics may be found in the substudy in Volume V.4.

Teacher Questionnaires

The 93-item teacher background questionnaire was developed by ETS and included some of the items used in the Coleman study. This was administered to all teachers in the fall. The questionnaire included:

1. demographic information: e.g., age, race, parents' socioeconomic status, community of origin;
2. teachers' preferences for and perceived effectiveness with student groups of differing racial, ability, or socioeconomic characteristics;
3. the presence and utilization of additional instructional personnel (paraprofessional or specialist); and
4. teaching as a career, subject matter taught, level and type of students currently taught, and school setting.

While it is impossible, due to space limitations, to present all of the responses to this questionnaire, items related to the four points above are included as examples in Table 17. The data in this table included two teachers who dropped out of the project; hence the percents are based on an N of 97 instead of 95.

A Teacher School Climate Questionnaire* was administered to all teachers in the spring. This instrument included items on organizational structure and teachers' attitudes. Responses to selected

*Developed by the staff of the Environments for Teaching Program, Elizabeth G. Cohen, Program Director. NIE Grant #NE-C-00-3-0062. Stanford University, Stanford, California.

Table 17

Selected Fall Teacher Questionnaire Responses

Demographic Characteristics	f	%
Age		
No response	4	4.1
20 - 29	18	18.4
30 - 39	26	26.8
40 - 49	28	28.9
50 - 59	19	19.6
60 - 69	2	2.1
Race		
Black	8	8.2
Chinese-American	1	1.0
Japanese-American	4	4.1
White	83	85.6
Other	1	1.0

Teaching Preference and Effectiveness	Prefer		Effective	
	f	%	f	%
Ability Level				
No response	2	2.06	2	2.06
A high ability group	18	18.56	19	19.59
An average ability group	36	37.11	27	27.84
A low ability group	3	3.09	5	5.15
A mixed ability group	32	32.99	27	27.84
More than one answer	6	6.19	17	17.53
Socio-economic Status				
No response	2	2.06	2	2.06
Mostly children of professional and white-collar workers	12	12.37	15	15.46
Children from a general cross-section of the community	76	78.35	60	61.86
Mostly children of factory, farm, and other blue-collar workers	--	--	2	2.06
Mostly children of unskilled workers	2	2.06	5	5.15
More than one answer	5	5.15	13	13.40
Racial or Ethnic Group				
No response	4	4.12	4	4.12
Predominantly white students	9	9.28	15	15.46
Mixed racial composition, but predominantly white	32	32.99	31	31.96
Mixed racial composition, but predominantly non-white	7	7.22	6	6.19
Predominantly non-white	2	2.06	3	3.09
Mixture of several groups, but no one predominant group	37	38.14	26	26.80
More than one answer	6	6.19	12	12.37

Table 17 (continued)

Selected Fall Teacher Questionnaire Responses

Presence and Utilization of Additional Personnel	f	%
Aides or other teachers		
Yes	73	75.3
No	24	24.7
Use of paraprofessionals or aides		
Yes	38	40.4
No	56	59.6
Number of paraprofessionals or aides used: N = 38		
No response	14	36.8
One aide	20	52.6
Two aides	3	7.9
Three aides	1	2.6
Use of other teachers		
Yes	23	24.5
No	74	75.5
Use of reading specialists		
No response	4	4.1
Yes	36	37.1
No	57	58.8

Number of children using reading specialists: N = 36	NR	Mean	Standard Deviation	Range
	26	5.28	5.92	1-25

Teaching as a Career	Second Grade f	Second Grade %	Fifth Grade f	Fifth Grade %	Total f	Total %
Satisfaction with job						
No response	--	--	4	7	4	4
Extremely satisfied	15	37	18	33	33	35
Very satisfied	18	44	23	43	41	43
Moderately satisfied	7	17	5	9	12	13
Slightly satisfied	1	2	3	6	4	4
Not at all satisfied	--	--	1	2	1	1

Ability level of students taught	f	%
No response	3	3.1
All high ability students	1	1.0
All average ability students	9	9.3
All low ability students	3	3.1
Combination of two ability groups	22	22.7
Heterogeneous grouping is used in this school	50	51.5
More than one answer	9	9.3

Table 17 (continued)

Selected Fall Teacher Questionnaire Responses

Subjects Taught	Second Grade f	Second Grade %	Fifth Grade f	Fifth Grade %	Total f	Total %
Mathematics						
No response	--	--	--	--	--	--
Yes	41	100	54	100	95	100
No	--	--	--	--	--	--
Reading						
No response	--	--	--	--	--	--
Yes	40	98	53	98	93	98
No	1	2	1	2	2	2
Social Studies						
No response	1	2	--	--	1	1
Yes	39	95	54	100	93	28
No	1	2	--	--	1	1

Grades Currently Taught	Second Grade f	Fifth Grade f	Total f
Grade 1	11	--	11
Grade 2	36	--	26
Grade 3	13	--	13
Grade 4	1	10	11
Grade 5	--	51	51
Grade 6	--	15	15

Classroom Setting	Second Grade f	Second Grade %	Fifth Grade f	Fifth Grade %	Total f	Total %
No response	--	--	1	2	1	1
Open space	6	15	2	4	8	8
Classrooms separated by movable partitions which are open most of the time	1	2	2	4	3	3
Classrooms separated by movable partitions which are closed most of time	1	2	--	--	1	1
Self-contained classroom	30	73	44	81	74	78
Other	3	7	5	9	8	8

items, later used to derive scores for variables included in the correlational, regression, and path analyses, are given in Table 18.

Likert Questionnaire

The Likert "Profile of a School"* questionnaire was administered to all teachers and principals as a measure of school climate. This instrument examines seven categories of organizational variables: motivation, communication, interaction, decision-making, goal setting, control, and performance. According to Likert, there is a relationship between school organizational climate and student achievement, even when achievement is corrected for student socioeconomic status.

Two scores were derived from the Likert teacher questionnaire for use in the path analysis. These are: 1) Principals's Administrative Style and 2) Class Climate. The items comprising the Principal's Administrative Style score included questions about friendly and supportive principal behavior, perceived confidence of principal solicitation of teacher ideas, perceived principal interest in teacher success, etc. The items which make up the Class Climate score include questions about teacher confidence and trust for students, perceived pupil trust for teacher, type and amount of interaction in class, teacher knowledge of student problems, perceived degree of student excitement about learning, etc.

The Likert principal questionnaire yielded two scores for use in the path analysis. These are: 1) Superintendent's Administrative Style and 2) Principal's (Perceived) Administrative Style.

---

*Profile of a School. Copyright © 1971, Rensis Likert Associates, Inc., Ann Arbor, Michigan.

Table 18

Selected Teacher School Climate Questionnaire* Responses

Teacher Characteristics	Second Grade f	Second Grade %	Fifth Grade f	Fifth Grade %	Total f	Total %
Number of years taught in current school						
No response	--	--	--	--	--	--
1-2 years	4	10	11	20	15	16
3-5 years	8	19	18	33	26	27
6-10 years	12	29	18	33	30	32
11-20 years	17	41	7	13	24	25
over 20 years	--	--	--	--	--	--
Total years of teaching experience						
No response	1	2	--	--	1	1
1-2 years	--	--	1	2	1	1
3-5 years	3	7	13	24	16	17
6-10 years	10	24	9	17	19	20
11-20 years	19	46	20	37	39	41
over 20 years	8	19	11	20	19	20

Organizational Structure	Second Grade f	Second Grade %	Fifth Grade f	Fifth Grade %	Total f	Total %
Grouping for instruction						
Mathematics - Most frequent grouping						
No response	4	10	6	11	10	11
Whole class grouped together	7	17	16	30	23	24
Two or three groups	15	37	14	26	29	31
Four or more groups	7	17	8	15	15	16
Students worked individually	8	19	10	19	18	19
Reading - Most frequent grouping						
No response	5	12	5	9	10	11
Whole class grouped together	--	--	3	6	3	3
Two or three groups	14	34	17	31	31	33
Four or more groups	16	39	16	30	32	34
Students worked individually	6	15	13	24	19	20
Social Studies - Most frequent grouping						
No response	9	22	7	13	16	17
Whole class grouped together	26	63	34	63	60	63
Two or three groups	3	7	3	6	6	6
Four or more groups	1	2	4	7	5	5
Students worked individually	2	5	6	11	8	8

Table 18 (continued)

Selected Teacher School Climate Questionnaire* Responses

Degree of Teaming	Second Grade f	Second Grade %	Fifth Grade f	Fifth Grade %	Total f	Total %
Which of the following most accurately applies to you?						
No response	1	2	1	2	2	2
I teach the same class throughout the whole school day	20	49	25	46	45	47
I teach the same students during most subjects, but have different students for a few subjects	14	34	20	37	34	36
I teach the same subjects to different classes	--	--	--	--	--	--
I teach different subjects to different sets of students	5	12	7	13	12	13
Other	1	2	1	2	2	2

Influence of school-wide policies on curriculum	Second Grade f	Second Grade %	Fifth Grade f	Fifth Grade %	Total f	Total %
No response	--	--	2	4	2	2
A great deal	8	19	6	11	14	15
A considerable amount	15	37	24	44	39	41
A moderate amount	15	37	12	22	27	28
Sightly	3	7	10	19	13	14
Not at all	--	--	--	--	--	--
Don't know	--	--	--	--	--	--

Helpfulness of principal for new teaching ideas	Second Grade f	Second Grade %	Fifth Grade f	Fifth Grade %	Total f	Total %
Providing teachers with ways to gain new teaching ideas						
No response	--	--	2	4	2	2
Extremely helpful	4	10	13	24	17	18
Very helpful	10	24	10	19	20	21
Moderately helpful	15	37	14	26	29	31
Slightly helpful	9	22	9	17	18	19
Not at all helpful	3	7	6	11	9	9
Backing up teachers on student discipline matters						
No response	--	--	1	2	1	1
Extremely helpful	12	29	24	44	36	38
Very helpful	13	32	9	17	22	23
Moderately helpful	11	27	6	11	17	18
Slightly helpful	2	5	8	15	10	11
Not at all helpful	3	7	6	11	9	9

*Developed by the staff of the Environments for Teaching Program, Elizabeth G. Cohen, Program Director. NIE Grant #NE-C-00-3-0062. Stanford University, Stanford, California.

Teacher's Expectations of Pupils

Teacher expectations for pupils are treated more extensively in a substudy presented in Volume V.2. Consequently, this instrument will be discussed very briefly here and the reader is referred to the substudy for more detailed information.

The basic theoretical question is, "Do teacher expectations produce differential academic growth?" If so, then it is desirable to ascertain how teacher expectations are communicated to students and what student characteristics bring about the different teacher expectations.

Each teacher was asked to rank all students in his or her class according to how well the teacher expected the student would perform the next year on standardized tests in reading and mathematics. These ranks were converted to deciles to adjust for class size.

Pupil Information from the Teacher

In the fall, teachers were asked to provide the following information for each child in their class: age, grade, sex, race, number of schools attended and for how long, and an estimate of the child's socioeconomic background. These data on the Pupil Survey Instrument are presented in Table 19. On a separate form, in the spring, the teachers were asked to indicate the child's native language and her or his attendance record. This information was used to drop records from the analysis of pupils who had been absent a good portion of the year.

Parent Questionnaire

A Parent Questionnaire, in both English and Spanish, was developed to obtain pupil background information not available in school records

Table 19

Pupil Survey Instrument

Year of Birth	N		60	61	62	63	64	65	66	67	NR
Second Grade	1052	f	-	-	-	-	40	260	681	69	2
		%	-	-	-	-	4	25	65	7	-
Fifth Grade	1608	f	3	50	426	1052	70	1	-	-	6
		%	-	3	27	65	4	-	-	-	-

Grade	N		1	2	3	4	5	6
	2658	f	75	864	117	70	1394	138
		%	3	33	4	3	52	5

Sex	N		Male	Female
Second Grade	1057	f	521	536
		%	49	51
Fifth Grade	1604	f	812	792
		%	51	49
Total	2661	f	1333	1328
		%	50	50

Racial/Ethnic Group	N		Black	Chinese	Filipino	Japanese	Mexican	Native American	White	Other
Second Grade	1052	f	116	29	25	65	208	4	553	52
		%	11	3	2	6	20	-	53	5
Fifth Grade	1603	f	210	22	41	67	261	1	902	99
		%	13	1	3	4	16	-	56	6
Total	2655	f	326	51	66	131	469	5	1455	151
		%	12	2	3	5	18	-	55	6

Handicapped	N		Yes	No
Second Grade	1026	f	79	947
		%	8	92
Fifth Grade	1504	f	90	1414
		%	6	94
Total	2530	f	169	2361
		%	7	93

Table 19 (continued)

Special School Experiences	N		Title 1 Reading	Bilingual Program	Remedial Reading	Miller-Unruh Reading	Follow Through	Head Start	Other
Second Grade	1062	f	6	26	53	71	3	8	273
		%	1	2	5	7	-	1	26
Fifth Grade	1618	f	103	23	85	97	6	15	279
		%	6	1	5	6	-	1	17
Total	2680	f	109	49	138	168	9	23	552
		%	4	2	5	6	-	1	21

Number of Schools Attended	N			NR	1	2	3	4	5	6	7	8	Out of Range
Second Grade	1058	K	f	3	901	48	2	-	-	-	-	-	104
			%	-	85	5	-	-	-	-	-	-	10
		1	f	1	787	215	19	4	-	-	-	-	32
			%	-	74	20	2	-	-	-	-	-	3
		2	f	66	636	274	48	8	2	-	-	-	24
			%	6	60	26	5	1	-	-	-	-	2
		3	f	941	74	28	8	6	-	-	-	-	-
			%	89	7	3	1	1	-	-	-	-	-
Fifth Grade	1618	K	f	3	1210	32	2	-	-	-	-	-	371
			%	-	75	2	-	-	-	-	-	-	23
		1	f	-	1148	197	14	4	1	-	-	-	254
			%	-	71	12	1	-	-	-	-	-	16
		2	f	-	1039	301	54	6	3	1	1	-	213
			%	-	64	19	3	-	-	-	-	-	13
		3	f	-	980	378	89	12	8	1	1	3	136
			%	-	61	24	5	1	1	-	-	-	8
		4	f	-	895	474	132	39	11	1	1	1	64
			%	-	55	29	8	2	1	-	-	-	4
		5	f	70	804	505	147	45	14	4	1	-	28
			%	4	50	31	9	3	1	-	-	-	2

Socio-Economic Background	N		Upper	Upper-middle	Middle	Lower-middle	Lower
Second Grade	1021	f	21	207	384	300	109
		%	2	20	38	29	11
Fifth Grade	1575	f	29	281	597	470	198
		%	2	18	38	30	13
Total	2596	f	50	488	981	770	307
		%	2	19	39	30	12

and to provide a check on the data obtained from other sources. This questionnaire was completed by approximately 70% of the parents. It provided information on pupil age, native language, school experience, television and reading habits, and also on parents' employment, education and training. Table 20, Parent Questionnaire, summarizes these data.

Principal Questionnaire and Interview

As part of the process of obtaining data on school climate, each principal completed a questionnaire* and took part in a structured interview.

The principal questionnaire was designed to provide descriptive information about the personnel of the school, the types of teaching arrangements utilized by teachers in the school, special funding received by the school, physical space in the school, number of pupils and record of absences at the school, and characteristics of the district itself. This questionnaire is concerned with the resources available to the school and their utilization. The questionnaire was developed at the Stanford Center for Research and Development in Teaching and was modified by ETS for this study.

The Principal Interview* was designed to assess the principal's perception of the following organization variables:

1) prevailing staffing patterns;

2) decision-making processes (who is involved in deciding what?);

---

*Developed by the staff of the Environments for Teaching Program, Elizabeth G. Cohen, Program Director. NIE Grant #NE-C-00-3-0062. Stanford University, Stanford, California.

Table 20

Parent Questionnaire

Year of Birth

	N		NR	60	61	62	63	64	65	66	67	68	69	73
Second Grade	743	f	8	-	-	-	-	26	179	465	60	1	2	2
		%	1	-	-	-	-	3	24	63	8	-	-	-
Fifth Grade	1204	f	27	3	33	299	784	55	1	-	-	-	-	2
		%	2	-	3	25	65	5	-	-	-	-	-	-
Total	1947	f	35	3	33	299	784	81	180	465	60	1	2	4
		%	2	-	2	15	40	4	9	24	3	-	-	-

Language

	N		NR	English	Spanish	Other
Second Grade	743	f	11	649	57	26
		%	1	87	8	3
Fifth Grade	1204	f	18	1041	93	52
		%	1	87	8	4
Total	1947	f	29	1690	150	78
		%	1	87	8	4

Special School Experiences

Child Participation	N		Title I Reading	Bilingual Program	Remedial Reading	Miller-Unruh Reading	Follow Through	Head Start	Other
Second Grade	743	f	61	39	30	23	19	60	107
		%	8	5	4	3	3	8	14
Fifth Grade	1204	f	133	70	77	28	49	109	184
		%	11	6	6	2	4	9	15
Total	1947	f	194	109	107	51	68	169	291
		%	10	6	5	3	3	9	15

Child Attended Nursery School or Pre-Kindergarten

	N		NR	Yes	No
Second Grade	743	f	12	444	287
		%	2	60	39
Fifth Grade	1204	f	24	534	646
		%	2	44	54
Total	1947	f	36	978	933
		%	2	50	48

Table 20 (continued)

Information About Parents

Years in School

	N		Did Not Attend	Grade 1-6	Grades 7-9	Grades 10-12	HS Graduate	Some College	College Graduate	Graduate School	NR
Second Grade	743										
Mothers		f	7	36	40	68	157	215	145	69	6
		%	1	5	5	9	21	29	19	9	1
Fathers		f	7	25	24	52	113	170	147	160	45
		%	1	3	3	7	15	23	20	21	6
Fifth Grade	1204										
Mothers		f	11	80	66	125	304	355	157	86	20
		%	1	7	5	10	25	29	13	7	2
Fathers		f	17	57	67	95	209	278	203	167	111
		%	1	5	6	8	17	23	17	14	9
Total	1947										
Mothers		f	18	116	106	193	461	570	302	155	26
		%	1	6	5	10	24	29	15	8	1
Fathers		f	24	82	91	147	322	448	350	327	156
		%	1	4	5	8	17	23	18	17	8

Employment and Training

	N		Employed Outside Of The House	Trained For An Occupation
Second Grade	743			
Mother		f	267	468
		%	36	63
Father		f	665	523
		%	89	70
Fifth Grade	1204			
Mother		f	533	708
		%	44	59
Father		f	1017	778
		%	85	65
Total	1947			
Mother		f	800	1176
		%	41	60
Father		f	1682	1301
		%	86	67

Table 20 (continued)

Total	N 1947	Employed As				Trained For			
		Mother		Father		Mother		Father	
		f	%	f	%	f	%	f	%
Workman or Laborer		24	1	96	5	4	-	17	1
Household Worker		2	-	-	-	-	-	-	-
Farm Worker		-	-	1	-	-	-	-	-
Farmer		-	-	-	-	-	-	1	-
Personel Service Worker, Fireman, or Policeman		84	4	80	4	93	5	44	2
Operator or "Semi-skilled Worker		55	3	144	7	24	1	70	4
Foreman or Craftsman		9	1	285	15	16	1	291	15
Office Worker		284	15	58	3	527	27	43	2
Salesman		24	1	62	3	12	1	49	3
Business Owner or Manager		48	3	290	15	15	1	93	5
Professional or Technician		259	13	551	28	467	24	643	33
Housewife		5	-	-	-	3	-	-	-
Don't Know or Could Not Assign		1153	59	379	19	786	40	696	36

Activities Done Most and Least with the Child

		Second Grade		Fifth Grade		Total	
	N	743		1204		1947	
		Most	Least	Most	Least	Most	Least
Read to Child	f	317	164	238	583	555	747
	%	43	22	20	48	29	38
Watch TV with Child	f	340	200	718	177	1058	377
	%	46	27	60	15	54	19
Play Active Sports with Child	f	196	322	358	446	554	768
	%	26	43	30	37	29	39
Play Quiet Games with Child	f	247	175	424	282	671	457
	%	33	24	35	23	35	23
Help Child with School Work	f	376	88	635	189	1011	277
	%	51	12	53	16	52	14

Table 20 (continued)

Child Does Spend Time Reading at Home

	Second Grade	Fifth Grade	Total
N	743	1204	1947
f	699	1088	1787
%	94	90	92

What Child Reads

	Second Grade		Fifth Grade		Total	
N	743		1204		1947	
	f	%	f	%	f	%
Family Books	558	75	807	67	1365	70
Library Books	491	66	822	68	1313	67
School Books	535	72	876	73	1411	73
Comics	155	21	353	29	508	26
Magazines	209	28	506	42	715	37
TV Guide	145	19	452	37	597	31
Newspapers	118	16	469	39	587	30

3) amount of influence the principal can exert in a number of different areas;

4) the characteristics of the school's reading program, grades 2 and 5;

5) teacher evaluation procedures;

6) available support staff (aides, reading or math specialists);

7) principal's background; and

8) principal's opinion regarding school climate instruments to be used with teachers.

The support for innovation score derived from these measures includes information on the number of extra teachers and aides (such as teachers who do not have a regular class, psychologists, counselors, speech therapists, librarians, curriculum specialists, reading specialists, and math specialists), the amount of district funds received, and the stimulation for change provided by the principal.

References

Bowles, S. S. Educational Production Functions, Final Report. Cooperative Research Contract OEC 1-7-000451-2651, 1968.

Bowles, S. S. and Levin, H. M. The determinants of scholastic achievement: An appraisal of some recent findings. Journal of Human Resources, 1968 3 (1).

Buros, O. K. Mental Measurement Yearbooks. Highland Park, N.J., The Gryphon Press. 1972.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "Structure of Intellect", Princeton, N.J., Educational Testing Service, RB-74-16, 1974.

Coleman, J. S. et al. Equality of Educational Opportunity. Washington, D.C.: Government Printing Office, 1966.

Commission for Teacher Preparation and Licensing, State of California. A report covering pre-conference planning and design conference results for the beginning teacher evaluation study. Sacramento, Author, 1973.

DiStefano, J. J Interpersonal perceptions of field-independent and field-dependent teachers and students. Unpublished doctoral dissertation, Cornell University, 1969.

Ekstrom, R. B. Cognitive factors: Some recent literature. Princeton, N.J., Educational Testing Service, PR-73-30, 1973.

Ekstrom, R. B. et al. Teacher performance and pupil growth. A report prepared for the California Commission Teacher Preparation and Licensing. Princeton, N.J., Educational Testing Service, 1973.

French, J. W., Ekstrom, R. B., and Price, L. A. Kit of Reference Tests for Cognitive Factors. Princeton, N.J.: Educational Testing Service, 1963.

Guilford, J. P. The Nature of Human Intelligence. New York: McGraw-Hill, 1967.

Hanushek, E. A. The education of negroes and whites. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1968.

Hanushek, E. A. The production of education, teacher quality, and efficiency. ERIC, 1970 (ED 037 396).

Hoepfner, R. et al. CSE elementary school test evaluations. Los Angeles: Center for the Study of Evaluation, University of California, 1970.

Hudson, L. Frames of Mind: Ability, Perception, and Self-Perception in the Arts and Sciences. London, England: Methuen, 1968.

Hunt, D. E., and Joyce, B. R. Teacher trainee personality and initial teaching style. American Educational Research Journal, 1967, 4, 253-259.

Lesser, G., Fifer, G., and Clock, D. H. Mental abilities of children from different social class and cultural groups. Monographs of the Society for Research in Child Development, 1965, 30 (4), No. 102.

Morsh, J. E., Burgess, G. G., and Smith, P. N. Student achievement as a measure of instructor effectiveness. San Antonio, Texas: Air Force Personnel and Training Research Center, 1955.

Murphy, R. T. Adult educational reading study. Princeton, N.J., Educational Testing Service, PR-75-2, 1975.

National Education Program Associates. Research design for phase II, Commission for Teacher Preparation and Licensing; beginning teacher evaluation study. Boston, Author, 1973.

National Education Program Associates. Review of literature for beginning teacher evaluation study. Boston, Author, 1973.

Quirck, T. J., Trismen, D. A., Weinberg, S. F., and Nalin, K. B. The classroom behavior of teachers and students during compensatory reading instruction. Princeton, N.J., Educational Testing Service, PR-74-5, 1973.

Rouse, W. M., Jr. A study of the correlation between the academic preparation of teachers of mathematics and the mathematics achievement of their students in kindergarten through grade eight. Unpublished doctoral dissertation, Michigan State University, 1968. University Microfilms No. 68-4205.

Taylor, C. W., Ghiselin, B. and Yagi, K. Exploratory Research on Communication Abilities and Creative Abilities. Salt Lake City, Utah: University of Utah, 1967. AFOSR 67-1523.

Thelen, H. A. Classroom Grouping for Teachability. New York: John Wiley, 1967.

Witkin, H. A. The role of cognitive style in academic performance and in teacher-student relations. Princeton, New Jersey, Educational Testing Service, RB-73-11, 1973.

Wu, J. J. Cognitive style and task performance: A study of student teachers. Unpublished doctoral dissertation, University of Minnesota, 1967. University Microfilms No. 68-7408.

APPENDIX

Representative Items from the Pupil Tests Developed for the BTES Study

Representative Items from the Pupil Tests
Developed for the BTES Study*

Attitude toward Reading and Mathematics

STUDENT ATTITUDE - SECOND GRADE*

1.	When the teacher asks me to think up some math problems, I feel	HAPPY	?	SAD
2.	When the teacher gives me a story that is too hard to read, I feel	HAPPY	?	SAD

.

.

STUDENT ATTITUDE - FIFTH GRADE*

1.	Reading is the hardest thing I have to do.	YES	?	NO
2.	I get nervous when the teacher asks me to do a math problem at the blackboard.	YES	?	NO

.

.

*All these tests are copyrighted by Educational Testing Service.

Pupil Survey

STUDENT EXPECTATIONS - SECOND GRADE*

Will you be a good math student next year?

_____ I will be very good in math.

_____ I will be O.K. in math.

_____ I will not be too good in math.

Here are some math problems. Put an X on the problems you think you can do. Do not try to solve the problems.

2 5 + 1	3 4 + 2	8 - 6
3 + 7	6 + 5	268 + 123

Will you be a good reading student next year?

_____ I will be very good in reading.

_____ I will be O.K. in reading.

_____ I will not be too good in reading.

Here are some things to read. Put an X on the ones you think you can read.

The cat is here.	Where is that silly alligator?	If we would be sincere, we should leave.

*All these tests are copyrighted by Educational Testing Service.

Application of Reading Skills

READING APPLICATIONS*

Directions: Put an X on the program about the San Diego Zoo.

2:00 p.m.	[4]	Baseball [Color] Teams to be announced. (Runs to 4:00)
	[6]	Top Cat - Cartoons
2:30 p.m.	[2]	Children's Variety [Repeat] Today the show goes to Detroit to watch cars being assembled (60 minutes)
	[6]	Visit the Zoo [Special] Famous San Diego Zoo is toured.
3:00 p.m.	[6]	Music Beat

Mark the letter of the coupon that shows you want to order both the pictures and the paints.

Picture and Paint Set
Box 2039
Battle Creek, Michigan 49015

A. Please send me one Cartoon Picture. Enclosed are 25 cents and one box top.

B. Please send me all four Cartoon Pictures. Enclosed are 50 cents and one box top.

C. Please send me all four Cartoon Pictures and the Paint Set. Enclosed are $1.00 and two box tops.

D. Please send me the Paint Set. Enclosed are 75 cents and one box top.

*All these tests are copyrighted by Educational Testing Service.

Decoding

DECODING: ROOT WORDS*

The teacher reads the following directions: "Look at the first word, the word with the line under it. It is the root word for one of the three other words in the line. Find the word that is made from the first word." The teacher then reads the root word. The child X's the word that he or she thought incorporated the root word.

Root Word	Choices		
<u>fly</u>	floating	flaps	flies
<u>near</u>	neater	need	nearer
<u>carry</u>	careful	carrying	carrot

Application of Mathematics Skills

MATHEMATICS APPLICATION*

Menu	
Soups	
Onion Soup	.60
Vegetable Soup	.75
Sandwiches	
Sliced Turkey	1.30
Roast Beef	1.50
Tuna Fish Salad	.95
Grilled Cheese	.75
Chicken Salad	1.10

Here is a menu. What would be the cost of your lunch if you had a cup of onion soup and a tuna fish salad sandwich?

$_____

*All these tests are copyrighted by Educational Testing Service.